Bansilal Ramnath Agarwal CharitableTrust's

Vishwakarma College of Arts, Commerce and Science, Pune.

(Affiliated to Savitribai Phule Pune University)

NAAC Accredited with 'B+' Grade

CRITERION III: Research, Innovations And Extension	
KEY INDICATOR	3.3 Research Publications and Awards
METRIC NUMBER	3.3.2

Number of papers published per teacher in the Journals notified on UGC website during the year

3.3.2.1 Number of research papers in the Journals notified on UGC CARE list year wise during the year

List of Research Publications: Academic Year 2022-23

Sr. No.	Title of the Paper	Name of the Author/s	Name of the Department	Name of the Journal	Page Number
1	An effective approach towards battery using Artificial Intelligence and Machine Learning	Dr.Sudhir Chitnis	Science	Empirical Economics Letters	1 to 49
2	An effective approach towards battery using Artificial Intelligence and Machine Learning	Sanket Lodha	Science	Empirical Economics Letters	
3	An effective approach towards battery using Artificial Intelligence and Machine Learning	Arun Patil	Science	Empirical Economics Letters	
4	An approach to enhance sentiment analysis on Social Media Data through Text Analytics and Predictive Modeling	Dr.Sudhir Chitnis	Science	Empirical Economics Letters	50 to 61
5	An approach to enhance sentiment analysis on Social Media Data through Text Analytics and Predictive Modeling	Arun Patil	Science	Empirical Economics Letters	
6	An approach to enhance sentiment analysis on Social Media Data through Text Analytics and Predictive Modeling	Sanket Lodha	Science	Empirical Economics Letters	

7	An approach to enhance sentiment analysis on Social Media Data through Text Analytics and Predictive Modeling	Prajkta Patil	**Science**	Empirical Economics Letters	
8	An approach to enhance sentiment analysis on Social Media Data through Text Analytics and Predictive Modeling	Swati Patil	**Science**	Empirical Economics Letters	
9	An Investigation into the Prognostic Indicators for Disease Exacerbations with Rheumatoid Arthritis Displaying Subdued Disease Activity: An Analytical Approach	Dr.Sudhir Chitnis	**Science**	Empirical Economics Letters	62 to 63
10	An Investigation into the Prognostic Indicators for Disease Exacerbations with Rheumatoid Arthritis Displaying Subdued Disease Activity: An Analytical Approach	Prajkta Patil	**Science**	Empirical Economics Letters	
11	An Investigation into the Prognostic Indicators for Disease Exacerbations with Rheumatoid Arthritis Displaying Subdued Disease Activity: An Analytical Approach	Swati Patil	**Science**	Empirical Economics Letters	
12	An Investigation into the Prognostic Indicators for Disease Exacerbations with Rheumatoid Arthritis Displaying Subdued Disease Activity: An Analytical Approach	Sanket Lodha	**Science**	Empirical Economics Letters	
13	Rural Non-Agricultural Entrepreneurship: Constraints	Dr.Shital Mantri	**Commerce**	Anvesak a bi-annual journal (UGC Care)	64 to 72
14	Rural Non-Agricultural Entrepreneurship: Constraints	Dr. Sheetal Waghmare	**Commerce**	Anvesak a bi-annual journal (UGC Care)	
15	Rural Non-Agricultural Entrepreneurship: Constraints	Vaishali Kale	**Science**	Anvesak a bi-annual journal (UGC Care)	
16	Issues and possibilities in marketing of agricultural commodities	Poonam Jadhav	**Commerce**	Anvesak a bi-annual journal (UGC Care)	73 to 78

R M Bhagat
Prepared by

IQAC Co-ordinator
Vishwakarma College of Arts Commerce and Science, Pune

Verified by

Empirical Economics Letters, 22 (2): (August 2023) *ISSN 1681 8997*

An effective approach towards battery usingArtificial Intelligence and Machine Learning

Authors

Aditya Patil[a], Vaishali Patil[a], Sanket Lodha[c], Sudhir Chitnis[b], Arun Patil[b]

[a]Vishwakarma Institute of Information Technology, Kondhwa, Pune- 411048,India
[b]Vishwakarma College of Arts, Commerce and Science, Kondhwa, Pune-411048,India
[c]MGM University, Institute of Management & Research MGM Campus, N-6, Cidco, Aurangabad, Maharashtra 431003, India

Abstract

Global warming and global pollution can consequence in many severe changes to the environment, ultimately challenging environmental problems and impacting human health. Human activities that result in an electronic-based population are the primary causes of the sharp rise in waste, particularly battery waste. Massive releases of heavy metals from battery waste impact health and ecosystems as a whole. Hence there is need to work on battery lifecycle and its affecting measuring factors. However, there are many challenges to enhance the lifecycle of battery and to sustain environmental balance.

Machine learning is a form of artificial intelligence that is fundamentally present in almost every area of our lives. It increased automation and increased productivity. This research article provides an overview of exemplary research efforts for efficient Machine Learning-based Battery Management Systems (BMS) to improve battery life cycles and address their measurement factors in the context of ecological development. This article covers thedesign and synthesis of battery materials along with recent advances in the field of machine learning Furthermore, this review article also covers future lines of potential research and development directions for Machine Learning techniques like clustering and regression in theprocess of determining battery lifetime.

Keywords: battery lifecycle; measuring factors; Artificial Intelligence; BMS; Machine Learning; RUL

* Corresponding Authors.

sdchitnis@vcacs.ac.in (Sudhir Chitnis), Vishwakarma College of Arts, Commerce and Science, VIIT Campus, Kondhwa (Bk), Pune, India.
principal@vcacs.ac.in (Arun Patil), Vishwakarma College of Arts, Commerce and Science, VIIT Campus, Kondhwa (Bk), Pune, India.

Introduction

The world's climate has deteriorated sharply in recent decades due to rapid increases in global temperatures caused by conventional energy sources and CO_2-emitting fossil fuel vehicles. It emphasizes the effects on food security, land and marine ecosystems, migration and displacement, and socioeconomic development. Energy storage system research and development have been initiated by this. Lithium-ion (Li-Ion) has emerged as one of the most relevant and widely used solutions in the development of energy storage devices, widely used in low-power portable electronic devices, electric vehicle power supplies, and grid-scale energy storage [1]. One of the most significant pieces of technology that has transformed the portable electronics and electric vehicle sectors today is Li-ion batteries. The most important characteristics of lithium-ion batteries include safety, long life, low self- discharge, high energy density, and high power density, and high-power density, which make it superior to other battery types like lead-acid and sodium sulfur (NaS) [2]. Lithium-ion batteries are the most widely used power batteries due to their obvious advantages of high voltage, low self-discharge rate, long life and high safety performance. Lithium-ion batteries are considered the most promising energy source due to their large capacity and high energy density [3]. The main power source of new energy vehicles, Li-ion Batteries, has a long historyof use in the automotive, aviation, and other industrial sectors [4]. Rechargeable battery use in electric vehicle applications has grown in popularity recently [5] Consequently, a recent study has revealed a 168% spike in Electric vehicles sales in India in 2021 compared to 2020 [6]. The adoption of electric and alternative fuel vehicles has undoubtedly risen due to increased understanding of their economic and environmental advantages, and this will also help the nation reach its goal of net zero emissions.

Although the energy density, power density, and C-rate of lithium-ion (Borah et al., 2020) batteries have improved significantly, nonlinear charge capacity and accelerated degradation rates remain concerns. Therefore, to overcome the inherent limitations of lithium-ion batteries, a self-adjusting Battery Management System (BMS) that accounts for measured factors such as sulfation and voltage limitation is required. To maximize battery performance, the correct operating state of the battery, including battery state of health (SOH) and state of charge (SOC), must be monitored and ensured [7-9]. For example, for electric vehicles, SOH is the odometer, and SOC is the fuel gauge for gasoline vehicles. Anticipating and working with measurement factors can help extend battery life. Moreover, accurate estimation of SOH and SOC can prevent battery overcharging and overheating, thus extending battery life [10]. However, it cannot be measured directly with electrical equipment. A controlled discharge test can be used to determine his SOC of the battery. The number of charge/discharge cycles

2

remaining or measurements of related physical parameters can be used to determine the SOH of the battery. SOH and SOC are predicted using information about the battery load current and terminal voltage. There are many methods of estimating SOC and SOH, each with its own strengths and weaknesses [11].

For large lithium batteries, knowing the SOC is extremely vital. Because lithium is the most chemically reactive of all commonly used cell chemistries, electronic battery management systems (BMS) is the only one required. One of the main tasks of BMS is SOC control [12]. Moreover, in automotive applications, one of the main uses of large lithium batteries, he needs to control the SOC very precisely in order to manage the energy flow effectively and safely. A battery's SOC indicates how much charge is currently available and varies according to the amount of load current drawn from the battery. Data-driven approaches can be further divided into two techniques: model-based methods and methods that measure SOC directly [13]. OCV-SOC has a flat curve and requires a relaxation time between two periodic measurements to get an accurate SOC [13]. OCV-SOC has a flat curve and it requires relaxation times to pass between two cyclic measurements in order to yield accurate SOC.

Adaptive techniques for SOC estimation include techniques such as Kalman Filter (KF), Fuzzy Logic (FL), Support Vector Machine (SVM), and Machine Learning [14–17]. Although the Kalman filter is the most well-known of them, this research article underlines the importance of Machine Learning methods.

1.1 Battery

A battery is a collection of electrochemical cells that power electrical equipment. Batteries continuously convert chemical energy into electrical energy and must be properly maintained for optimum. A soiled battery may discharge through the muck on its top case.

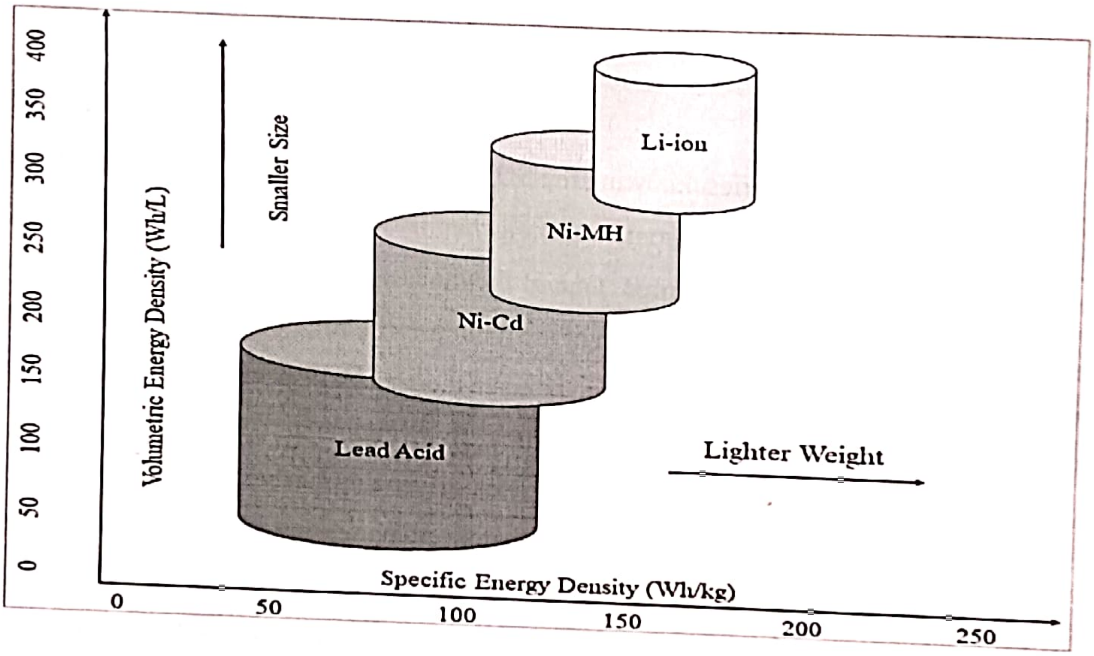

Figure 1: Different types of battery showing Li-ion batteries are ahead of most other battery types in these respects [Roberta Dileo, Rochester Institute of Technology- Clean Energy Institute].

Electrolyte, negative and positive electrodes, and battery case make up a battery. At the two spatially separated electrodes, chemical processes occur, and an external conducting circuit connecting the electrodes can provide electrical power. Electrolytes don't have much electronic transport and conduct electricity by the movement of ions. Ions and electrons are both carried by electrodes. The batteries can be raised in voltage to achieve connected in series with others, and it is required to link them in parallel to increase capacity. Primary batteries, rechargeable batteries, and fuel cells are the three main types of batteries [18]. One or both of the electrode reactions in a main battery are irreversible, making it impossible to recharge the battery.

The electrochemical reactions in a secondary battery can be turned around, allowing the battery to be recharged hundreds or even thousands of times. Reactants are continuously delivered into a fuel cell from the outside.

In addition to battery usage, we use a management system with special monitoring features such as charge management mechanisms and temperature control to prevent health, safety and property risks. These systems use merit-based metrics to control battery performance. The performance and condition of the battery are estimated using data such the

4

state-of-health (SOH) and state-of-charge (SOC) [19]. In this review article, we suggest a clever way to research the above-mentioned factors utilizing a data-driven methodology. To estimate these values, we propose a machine learning method that separates the important features from the discharge curve. Knowledge can be extracted from complex data without relying on previous underlying relationships between data variables [20]. Numerous simulations have been proposed to assess how well the suggested approach performs in various currents and temperatures.

1.2 What is AI and ML?

Machine learning is an application of artificial intelligence. Machine learning enables a wide variety of software applications to predict outputs without being explicitly programmed. Machine learning algorithms build models based on sample data as training data and can make predictions and decisions without being explicitly programmed [21-22]. AI is central to the development of modern robotics [23] autonomous driving [24] and smart grids [25]. Artificial Intelligence (AI) and its branch of production, known as Machine Learning (ML), is an excitingprospect that will bring about major changes in battery research and development [26] and how to overcome major issues while dealing with large number of data and factors.

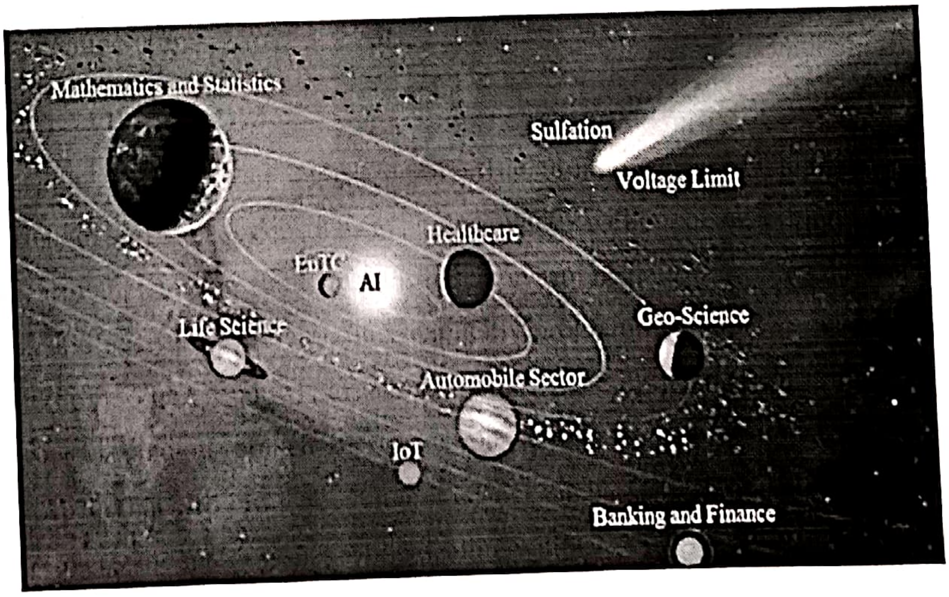

Figure 2: AI and the various areas revolving around them and affecting measuring factors

5

Machine Learning algorithms are used in many applications, such as computer vision and email filtering, where traditional algorithms fail to do what is expected. Computers learn from the information you give them to perform specific tasks. Machine Learning evolved basedon the ability to use computers to explore the structure of data, but we have no theory of what that structure should look like. Machine learning model testing is new data validation error. They often use an iterative approach to learn from data, and that learning can be easily automated.

The following Figure 3 illustrates how Machine Learning for Remaining Useful Life (RUL) estimate has been deployed geographically by different nations. The number of papers that have been evaluated makes it evident that Machine Learning techniques for RUL estimate are more frequently recommended expansion in India (represented in green color) by taking into consideration their implementation in China and other countries.

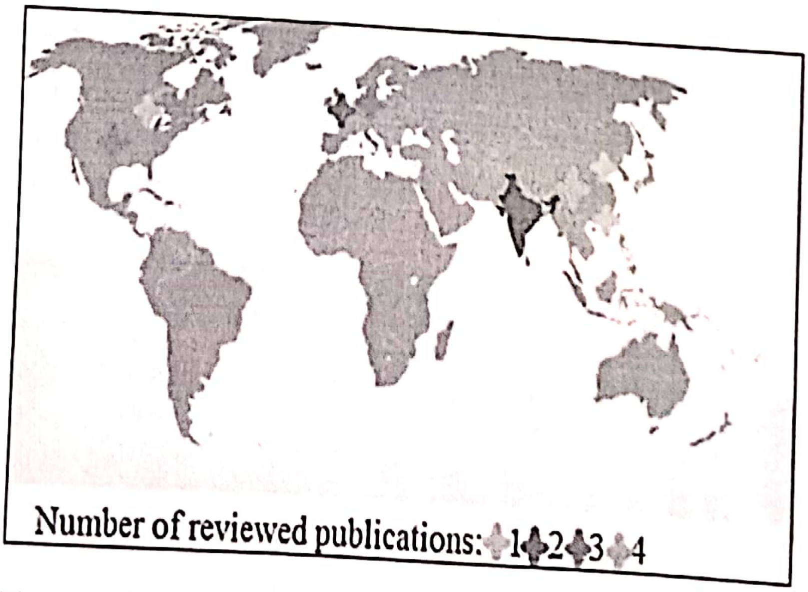

Figure 3: Research in Machine Learning RUL estimation by country

New approaches to predict battery degradation include supervised and unsupervised algorithms such as probability estimation such as time series analysis, SVM, ANN, and regression. The most common are based on probabilistic estimation.

1.2.1 Introduction to Machine Learning

Machine learning (ML) algorithms can extract knowledge from very large and complex data sets without relying on previous underlying relationships between data variables. This study provides insight on the Machine Learning domain and its methodology for analyzing and forecasting battery-affecting factors. However, Machine Learning models are built on

6

Statistical concepts. They are classified into four learning theories: regression, rule extraction, clustering, and classification, each of those including several algorithms [27]. In this article, we are emphasizing Machine Learning paradigms also known as Supervised Predictive Learning Models (SPLM). Starting with very simple regression techniques, we applied more advanced clustering and classification algorithms to validate battery-related issues.

In order to predict the potentially non-linear correlations between the most important battery-related elements, various SPLM are being investigated more and more. The benefit of SPLM is that they are more flexible and make fewer, if any, assumptions about the input variables, making them better equipped to interpret outliers, noisy data, or missing data. Battery issues are analyzed using mathematical and/or statistical models and predictions are made after a general understanding of the data. The main difference between SPLM and statistical models is that SPLM aims to understand the structure of data and fit theoretical distributions to well-understood facts. Statistical models, by contrast, have mathematically backed theories, but also require data that adhere to a rigorous set of assumptions.

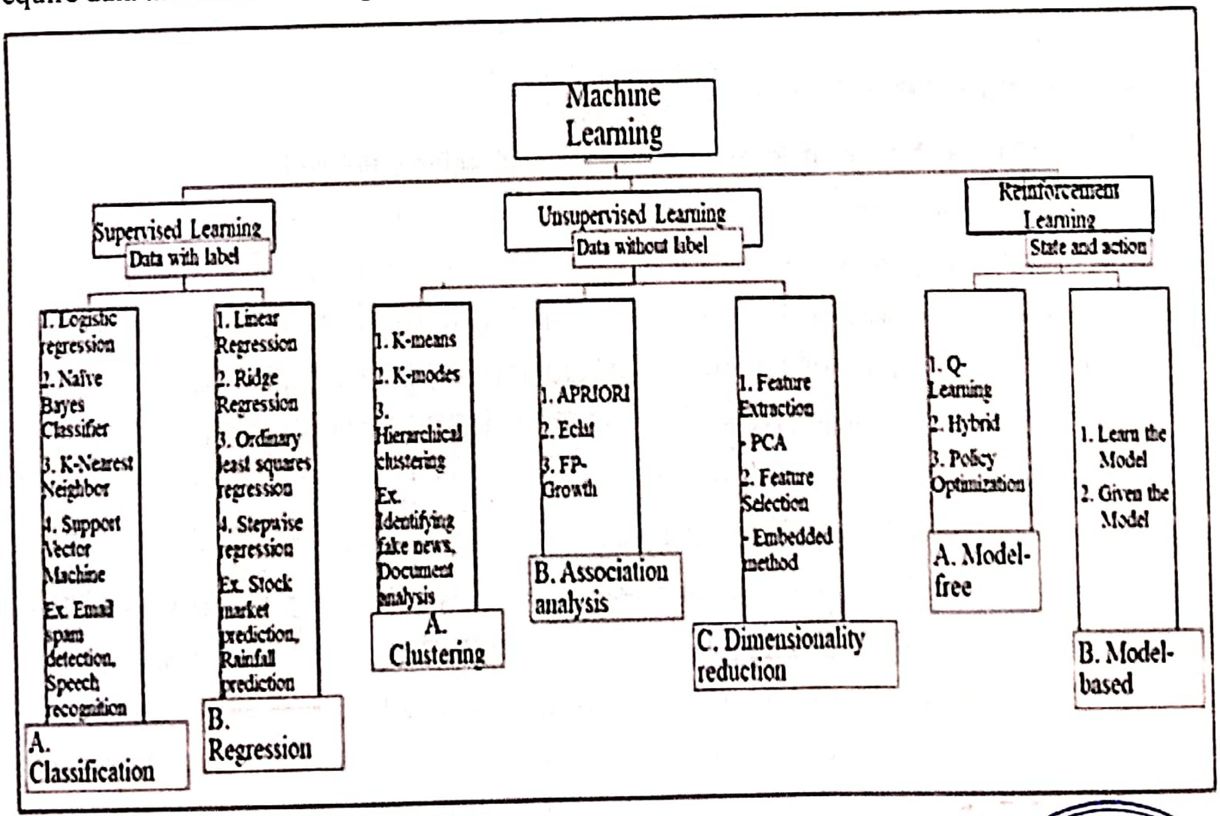

Figure 4: Machine Learning Algorithm Classification

Considering the employment of sensors in future work, SPLM was chosen for this research study not only for its effective data handling, but also for the algorithm selection and data collecting quality. Additionally, it can be applied to applications that include pattern recognition and prediction. According on the type of the provided input, ML techniques are typically split into threediverse categories as follows:

7

1.2.1.1 Supervised Learning

It is a method of completing a task by giving the systems training, input, and output patterns. Figure 4 shows supervised learning algorithms that were taught using labelled data. In supervised learning, the model receives input data in addition to output. In this kind of learning, sample inputs and expected outputs are sent to the computer. Learning a general rule that connects inputs and outputs is the objective. It means to provide concept characteristics or properties, such as sulfation and voltage restrictions, and to create our own models to recognize items with comparable structural traits. Classification, decision-making, and regression are some examples of supervised learning.

Here the algorithm learns by mapping the actual output to the correct output and detecting errors. Then modify the model accordingly. This work describes regression and other methods that use clustering models to predict label values for additional unlabeled data. Supervised learning is commonly used in applications where historical data predicts likely future events.

1.2.1.2 Unsupervised Learning

This is a self-directed learning technique that requires the system to determine the characteristics of the input population itself and does not use a previous set of categories. In unsupervised learning, only input data is provided to the model. In this type of learning, the learning algorithm is not given a label, so the learning algorithm itself can find the structure of the input. Unsupervised learning can be a goal in itself by discovering hidden patterns in data. That is, build models from classes that are not predefined. Examples: clustering and dimensionality reduction.

In unsupervised learning, a machine cannot be taught the "correct answer". This works well with transactional data. For example, the study can identify battery factors with similar properties, such as passivation and corrosion, and predict similarly for battery types. However, we identify the main factors such as sulfation, voltage limitation from each other.

Examples of unsupervised techniques include nearest neighbor mapping, self-organizing mapping, singular value decomposition, and k-means clustering. All of these algorithms are used for article recommendation, text topic segmentation, and data outlier identification.

8

1.2.1.3 Reinforcement Learning

In this type of learning, a computer program interacts with an active environment in which certain goals must be achieved. As you traverse, the computer program gives you reward-like feedback. The program also tries to maximize the rewards given.

1.2.2 Artificial Neural Network

The functioning of biological neurons acts as the inspiration for the Artificial Neural Network (ANN). They are computing models originated by an animal's central nervous systems. ANNs, one type of ML model, nowadays are competing with traditional statistical and regression models in terms of utility [28]. It is a processing architecture based on how the human brain presents information and is able to learn and adapt. They are mainly used to address classification, prediction, classification, and optimization problems. Mathematical representations of biological neurons, known as perceptron, are grouped into nodes and connected by weight vectors (simply weights) to form ANNs. By continuously adjusting the weights between nodes based on the flow of information through the network throughout the training phase, ANNs can model all real-world data variability. ANNs are ideal for modeling complex interactions between inputs and outputs, and their ability to learn and adapt makes them particularly effective tools for modeling nonlinear statistical data.

To model the Li-ion battery effectively and online in order to estimate the State of Charge (SoC). Experimental data can be used to train both artificial neural networks (ANNs) in offline mode.To predict SoC, a Feed Forward Neural Networks (FFNN) network uses a nonlinear autoregressive model with an eXogenous Input (NARX) network to determine the required battery voltage. Simulation results for this method show good accuracy and fast convergence.

Estimating battery State of Charge (SoC) is a difficult process that requires inference rather than direct measurement. Most applications today use temperature monitoring along with voltage and current measurements at the battery terminals to view the battery from the outside. These signals can be used to estimate his SoC using various existing approaches, as described in recent publications [29-32]. Artificial Neural Networks (ANNs) have been investigated in many research projects to connect acoustic signatures to SoCs in bicycle bag cells for SoC estimation. Some research has shown that the process of training and testing regression models is streamlined because the full waveform can be processed without pre-selecting features.

Empirical Economics Letters, 22 (2): (August 2023) *ISSN 1681 8997*

It is also demonstrated that simple feature filtering techniques based on statistical significance can be applied, yielding SoC predictions comparable to or better than those obtained with the full waveform. This is done considering the motivation behind feature selection and dimensionality reduction in the regression process. Statistical significance is calculated using the Pearson correlation coefficient between any feature and SoC. Many studies have used a subset of Fourier coefficients in the frequency domain to assess their predictive power as regression functions in the time domain.

By applying a feed-forward neural network regression model, acoustic signatures generated experimentally from a 210 mAh LCO ($LiCoO_2$) pouch cell during CC-CV (Constant Current-Constant Voltage) cycling have been used to carry out SoC inference [33]. Elias Galionas et al. The following different data structures were considered: the complete signal waveform, waveform features selected based on the Pearson correlation threshold for SoC, and the lowest frequency Fourier coefficient magnitude of 15%, the first three types of waveform information.

This paper states that a moderately sized FFNN can be perfectly trained on acoustic waveforms without the need for feature selection. A mean absolute estimation error close to 1% was achieved with this method, demonstrating its potential for use in a simplified workflow. While using a reduced input vector of time-domain features chosen based on statistical significance, the SoC estimates obtained by the filtered data construction are of comparable accuracy. They provide an effective, formal, and reliable dimensionality reduction method for applications where computational economy is paramount. The frequency domain setup offers the best accuracy with a mean absolute error of 0.75%. Importantly, all data structures are generated with very high accuracy using only acoustic information without the aid of voltage or temperature measurements. This demonstrates the potential of the acoustic method for use as a standalone SoC estimation method in operando with obvious security and reliability advantages. There are also new applications for failure detection of corresponding voltage sensors and online evaluation of voltage SoC mapping by comparison with acoustic SoC estimation.

Furthermore, method which was analyzed in this study and proposed by Elias Galiounas and et al. tested on data sets containing numerous cells, where bigger changes in operation temperature and innate electrochemical stiffness may provide additional difficulties. Hence, the research article states that machine learning strategies using FFNNs are likely to be the most appropriate.

[7] Kanakaraj, M., Guddeti, R. (2015) Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques", Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), IEEE ICSC 2015 February 7-9,2015, Anaheim, California, ©2015 IEEE

[8] Fersini, E. (2017). Sentiment Analysis in Social Networks: A Machine Learning Perspective. In University of Milano-Bicocca (Chapter 6). Milan, Italy, Sentiment Analysis in Social Networks. http://dx.doi.org/10.1016/B978-0-12-804412-4.00006-1

[9] Pathak, R., & Waghmare, G. (2020). Remarketing through Social Messaging Apps- Promotional Tool or Sales Strategy. Test Engineering and Management, 83, 14088 – 14099.

16

Hybrid techniques have recently been developed to increase estimation accuracy. Wei [34]developed a hybrid technique based on ANN and UKF for SoC estimate. Based on the current, voltage, and temperature measurements performed by the ANN, the state of charge ofthe SoC can be calculated. To reduce the ANN error, we can apply an unscented Kalman filter. [35] performed SoC estimation using Radial Basic Neural Network and Extended Kalman Filter (EKF). The combined model has the best performance in terms of estimation accuracy with an error of less than 1%, but the analytical linearization of the EKF introduces numericalproblems in the Jacobian computation of the model. As a result, this work proposed to treat thebattery model as a black box and use a NARX model (a nonlinear autoregressive model with eXogenous inputs). Then compute the SoC using a second feed forward neural network.

New battery models for SoC estimation have been proposed in several articles. Some researchers use dynamic battery models to predict his SoC. W. He's work [34] considers the effect of temperature and his SoC on the battery model. The battery was charged from 0% to 100% and discharged from 100% to 0%. Therefore, the solid curve is taken to represent the experimental His SoC. The integration error was minimal if the current sensor was accurately calibrated.

The paper proposed an ANN-based SoC estimation system for lithium-ion batteries. ANN is of type FFNN and NARX. NARX was trained offline to find the appropriate model needed by the FFNN to compute the SoC of the battery. The entire experimental database for this study was obtained from the NASA Prognostic Center of Excellence website. Simulation results of the proposed estimator showed good accuracy and fast convergence to the experimental variables, regardless of the loading conditions. The proposed model can be applied to many secondary batteries.

By following above, it is essential to explore a few issues with the suggested model. Calculating SoC for each cell is important to apply this approach to a battery pack. Batteries are employed in a variety of environmental settings. As a result, the database utilized to create any model must consider to every conceivable operation scenario.

1.3 Materials Design and Synthesis

The design and synthesis of battery materials are combined with recent advances in machine learning. Section 1.3.1 first provides a brief overview of recent initiatives to develop suitable material descriptors. This typically represents the first hurdle to implementing a practical and accurate machine learning model. We then present some examples of how machine learning-based research is accelerating the screening and prediction of new battery materials with specific desirable properties.

Section 1.3.2 discusses how machine learning algorithms can help handle more complex chemistries, longer length and time scales, and multiscale modeling, thereby opening up new possibilities for materials simulation. Subsection 1.3.3 discusses the creation of new materials, especially those arising from using artificial intelligence to efficiently organize experiments and exhaustively render chemical and physical spaces in conventional spaces with high-throughput methods. Describes how to solve the combinatorial explosion problem. (HT). In this context, we show how ML algorithms can be used to discover relationships between variables and predict outcomes for new tests. In subsection 1.3.4, we present our thoughts on applying artificial intelligence/machine learning to materials design and synthesis while outlining some of the major hurdles ahead.

1.3.1. Discovery of Materials

Finding new active electrode and electrolyte materials for next-generation batteries is the goal of current technology in this regard, often combining HT screening with machine learning. Typical machine learning techniques include decision trees, support vector machines, and artificial neural networks.

Machine learning approaches are used to discover complex nonlinear correlations between large numbers of variables based on the use of high-fidelity data. This data comes from physics-based simulations, experiments, or both. Ultimately, this helps classify materials with comparable properties or predict desirable properties for new materials.

Properties like discharge capacity, volume change, capacity retention, voltage profile, etc. are frequently of interest in the case of active electrode materials. The current invention relates to active electrode materials and manufacturing processes for them. These substances are valuable as active electrode components for Li-ion or sodium-ion batteries, such as anode components for Li-ion batteries [36]. The search for inorganic solid ionic conductors with high ionic conductivity and favorable mechanical properties is currently the focus of electrolytes.

12

The first obvious obstacle to establishing any Machine Learning model is the lack of sufficient, trustworthy training data. Training a model with the data that is already available and evaluating its classification performance using newly gathered data or a different dataset are both reliable methods for validating the performance of ML models [37]. Data quality is important, and dataset size has a large impact on accuracy. It's also important to thoroughly sanitize your data. Also, this data will not cause any errors during the training process. Errors can affect model performance later. Therefore, curation is essential and a must for building an accurate database that is as objective as feasible, regardless of the data source [38-39]. High-quality data implies that the simulation must successfully replicate the goal property in the real world (high fidelity). High-fidelity data, however, is typically hard to come by because it is expensive to acquire. Moreover, despite the tendency of scientists to primarily report on the best materials in the literature, it is common practice to enter unconvincing materials or failed experiments into databases in order to improve training sets. requirements. The underlying mathematical description of different compounds must be accurate enough to allow comparison of different structures and chemistries across huge data sets, another of the machine learning approaches applied to materials science. An important factor. The difficulties that arise when big data and Machine Learning are combined with materials science have received a lot of attention [40]. The choice of acceptable descriptors that result in sufficiently exact predictions of the intended target property is a crucial hurdle when using Machine Learning models to a materials science problem [41]. In reality, a given ML model's prediction performance is frequently decreased by subpar descriptions unrelated to the target features.

It is also crucial to note that the field of Artificial Intelligence-aided materials discovery is expanding quickly and has a number of promising future research directions. Inverse design is one of them, and it helps hasten the search for ultra-high-performance batteries. A major European research initiative, Battery 2030+, recently identified the establishment of the Battery Interface Genome (BIG) and Materials Acceleration Platform (MAP) as key milestones to accelerate the discovery of ultra-high performance batteries. bottom [44]. Initiatives in this direction include the MAP and the so-called BIG [42]. A deep generative model is essentially a generative model specified as a deep neural network. The purpose of generative modeling is to learn the underlying data distribution [43]. BIG-MAP focuses on the use of deep generative models that can generate new data with target characteristics, especially high- power output of batteries by merging data from different experimental approaches and simulation methods. Allows reverse design of the mesophase.

1.3.2. Active Electrode Materials

Input factors include firing temperature, Ni content, primary particle size, coating material, and cleaning conditions. The initial variables were initial capacity, cycle life, and amount of Li remaining after synthesis. According to the authors' evaluation of various machine learning models (ANN, RF, LR, k-NN, SVM, and DT), the highly randomized tree (ERT) has the lowest error in the projection of all output variables. generated. In a recent study, Kireeva and Pervov used his SVM model to find correlations between the synthesis and electrochemical properties of Li-rich layered oxide cathodes [45]. The output variables in this case were initial discharge capacity and coulombic efficiency, and the input factors in this case were composition, synthesis process, Li and transition metal sources, Li excess, temperature, calcination and sintering time. Key factors affecting custom properties such as: B. Specific processing settings, Li excess, or Li to transition metal ratios can be identified by machine learning analysis.

In order to avoid unworkable Edisonian-style experiments that aim to cut down on human effort and research expenditure, machine learning (ML) has received enormous interest [46]. For instance, Sendek et al. [47] screened approximately 12,000 innovative solid Li superionic conductors for all-solid-state LiB using logistic regression.

Wang et al. for the purpose of designing stress-relaxed structures, lattice constants calculated using density functional theory (DFT) of fully lithiated and delithiated structures were used to study spinels and layered oxides. A partial least squares (PLS) model was developed to predict the volume change rate [48]. They found that the radius of the transition metal ion and the distortion of the transition metal octahedron are the most important variables for accurately predicting the volume change. Joshi et al. Focused on predicting voltage profiles for a wide range of active electrode materials for Li, Mg, Ca, Al, Zn, and Y-ion batteries using DFT-calculated voltages contained in the Materials Project Database (MP) [49]. An asymmetrical three-dimensional structural arrangement of ions, atoms, and molecules to form a unit cell is called a crystal lattice. The type of crystal lattice is determined by the substantial and distinctive geometrical shape of a unit cell. The type of crystal lattice and other elemental characteristics of the atomic constituents in each specific compound

14

were among the descriptors of the materials that were taken into consideration. Some non-toxic and relatively common in comparison to lithium, multivalent and monovalent metals, including as magnesium, zinc, and aluminium, are thought of as substitute guest ions for Li-ion, leading to the development of several new battery systems. Their inability to accommodate alkalis with large radius (Na^+, K^+) or multivalent ions with strong electrostatic repulsions (Mg^{2+}, Zn^{2+}, Al^{3+}) has been the main barrier to their wide-scale deployment [50]. Considering known Li-based active electrode materials that have not yet been proposed for other chemistries, this work helped identify potential new electrode materials for Na and K ion batteries.

To support the search for organic electrode materials, Allam et al. created a DFT-based database of a compiled collection of various organic compounds [51]. In his ANN-based prediction of redox potential, the author considered both calculated electronic properties and optimized geometric data as input variables. The following 10 main input variables were discovered by the authors using linear correlation analysis based on calculation of Pearson correlation coefficients: electron affinity, highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), HOMO-LUMO gap, number of H, C, B, O, Li atoms and aromatic rings in the molecule. Examining compounds that were not part of the training set, the method showed excellent ability to predict accurate redox potentials with respect to DFT calculations.

Additionally, machine learning can be used to classify datasets into specific classifications using supervised learning. Logistic regression is a supervised machine learning algorithm used for classification purposes [52]. Here, Attarian Shandiz and et al. analyzed data from the Materials Project and a number of methods like ANN, SVM, k-NN, RF, ERT and LDA to determine Crystal systems (monoclinic, orthorhombic, and triclinic) of lithium-ion silicate-based cathodes containing Mn, Fe, and Co [53].

From above, the authors conclude that crystal volume, number of sites, formation energy, over-cladding energy, and bandgap are the most important descriptors, with HF and ERT classifiers producing the lowest overall errors. showed.

1.3.3. Solid Electrolytes

Solid electrolyte needs excellent ionic conductivity as well as chemical, thermal, and electrochemical stability in order to be used in practical applications [54]. The research paper,

[18] by Arun Patil and et al. reviewed recent material developments of fast solid Li-ion conductors. The paper states that glassy solid electrolytes have great potential for various electrochemical applications, the most important of which is solid-state batteries. Especially in the field of solid electrolytes, screening of battery materials using machine learning models is active. One of his earliest research projects proposed a new olivine-type oxide solid electrolyte with low ionic conductivity in 2012 by combining computer data with his PLS analysis [55]. In particular, scientists used the nudged Elastic Band (NEB) approach to calculate the DFT Li+ transfer energies for various ordered LiMXO4 structures within the M2+X5+ and M3+X4+ main group pairs. As a result of predicting the existence of materials with Li+ migration energies below 0.3 eV, several innovative compositions have been proposed including MgAs, ScGe, InGe, and MgP as well as AlX, GaX, InX, and CaX pairs more generally. Jalem et al. also explored tavorite-type LiMTO4F (with M3+T5+ and M2+T6+ pairings) solid electrolytes using an ANN model as an alternative [56] Low migration energies were predicted for certain compositions, such as LiMgSeO4F, LiMgSO4F, or LiGaPO4F. Scientists have sometimes used bond valence field (BVFF)-based calculations rather than DFT to create computational databases. BVFF-based computations are significantly less computationally intensive than DFT, making it easier to filter large amounts of data at once [57]. In contrast, Kireeva and Pervov took into account a wide range of experimental information about garnet-type oxide solid electrolytes.

The authors focused on predicting the ionic conductivity of compounds of general formula A3B2 (XO4)3 using the SVM model. Fujimura et al. used the SVM approach to directly evaluate the ionic conductivity of Li8cAaBbO4 LISICONs by merging theoretical and experimental data sets [58].

The desired property for each of the research previously mentioned is ion mobility (migration energies or conductivities). Among the reported accidents, an internal short circuit due to lithium dendrite formation was one of the causes of battery failure [59]. However, the prevention of dendrite formation must also be considered when choosing a suitable solid electrolyte. Ahmad et al. performed a DFT-based computer-aided screen to evaluate the potential of thousands of inorganic solids to suppress dendrite initiation when in contact with lithium metal [60]. To achieve this, the authors used the calculated shear and mass modulus and elastic constants to train CGCNN (Crystal Graph Convolutional Neural Network) and KRR models. They found that soft and highly anisotropic materials with high mass densities,

16

Li ratios and sublattice-bound ionicity are the most effective ways to inhibit dendrite formation.

1.3.4. Liquid Electrolytes

All-solid-state lithium batteries have advantages such as flexibility in cell design, resistance to electrolyte leakage, and high safety. As a result, some efforts have been made so far to increase the ionic conductivity of solids, especially considering their potential use under ambient conditions [1 x 10-3 Scm-1]. [18].

As liquid electrolytes are far more disordered than solid electrolytes, it is more difficult to do Machine Learning studies of energies and electrical and structural characteristics. As a result, there are a lot less studies that use Machine Learning techniques for liquid electrolytes. However, there are a number of uses for Machine Learning approaches that are applicable to liquid electrolytes, and we highlight a few of these uses in the following. Nakayama et al. [61] used an exhaustive search by GP (ES-GP) approach to predict cation-solvent interaction energies in liquid electrolytes of lithium-ion batteries, considering a variety of commercially available organic solvent compounds. Since the solvation and desolvation of Li ions at the electrolyte-electrode interface is a key process that often limits the overall mass motion, the interaction energy is a useful predictor of Li-ion transport within the electrolyte.

Furthermore Sodeyama et al., An exhaustive search using a linear regression model (ES-LiR) was used to add melting point as a target attribute. This is an important variable related to the wide LIB working temperature window [62]. Compared to multiple linear regression (MLR) and least absolute shrinkage and selection operator (LASSO) approaches, the ES-LiR model offered a fair compromise between predictive accuracy and computational complexity. However, the ES-GP method is much more accurate than his ES-LiR [63].

Utilizing Machine Learning techniques to speed up the exploration of the sampling space is a less direct strategy. For instance, ML-enhanced MD simulations can simulate more extreme varieties of liquid electrolytes. When dipole polarization is a key factor in the dynamics, for example, Ionic-liquid-based electrolytes (IL) and highly concentrated electrolytes (HCE). The most straightforward method is to use NNs, such as those in references [64-65], a cost-effective way to learn the concept of polarization. However, with the caveat that details of physical interactions are lost, this provides a productive techniques for data generation. Extensive research has led to the development of polymer electrolytes for Zn-based flexible batteries [66]. A similar method has been used to study his Zn2+ electrolyte in water, showing that ANNs can be applied to obtain useful physical potentials even in highly disordered systems [67].

Even though it is all too frequently overlooked, Machine Learning can be used in electrolyte investigations from both an experimental and computational standpoint. ML (Machine Learning) can be utilized in a variety of ways to improve experimental research of electrolytes, from completely automated laboratories to using ANN to evaluate spectroscopy data [68-70]. This method addresses several issues with sophisticated electrolyte simulation and comparisons with experimental data, allowing for more effective experimental research of liquid electrolytes.

1.4. Battery Measuring Factors

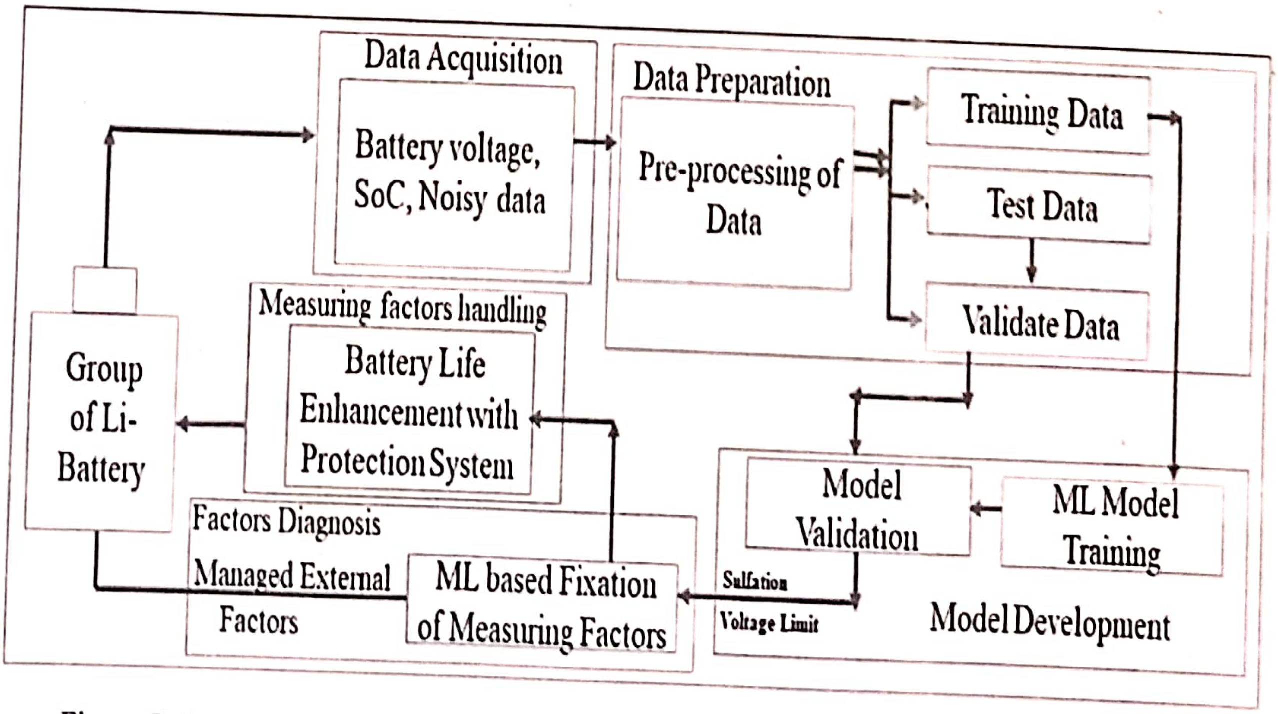

Figure 5: Basic block-diagram of ML-based life enhancement and protection system

The battery life enhancement and protection system of Li-ion batteries are imperative in Battery Management System. Therefore, detection of measuring factors such as sulfation, voltage limit, etc. in Li-ion battery and designing an efficient advance defense system are of utmost importance. Naturally, to establish an appropriate battery life enhancement and protection policy, in-depth knowledge and understanding of different kinds of factors alongside the identification of corresponding measuring factors mechanism in Li-ion batteries are highly essential. Hence, this review article focuses on the same.

18

1.4.1 Sulphation

Sulfation can permanently reduce the capacity of lead-acid batteries if they are kept in a low state of charge (SOC) for long periods of time without recharging [71].

Sulfation occurs and builds up when the battery is not fully charged. It remains on the battery plate [72]. Exceeding sulfation interferes with the chemical-to-electrical conversion and can seriously affect battery performance. Sulphation is normal for starter batteries in cars driving in congested areas. This can also occur when the engine is idling or running at a low speed and the battery cannot be fully charged.

Sulfation is a major problem in lead-acid batteries. Therefore, desulfurization is the solution for restoring sulfate lead-acid batteries. However, de-sulfation was not found as a better solution to prevent sulfation. Therefore, a battery management system for proper charging or discharging is found as a passive solution to sulfation. There are many methods in the literaturerelated to proper charge and discharge control that may involve sulfation issues. Therefore, in this article, an Atom Search Algorithm (ASA)-based Hybrid Energy Storage System (HESS) was developed to enable proper charge-discharge control to extend the life of lead-acid batteries by avoiding sulfation [73]. Therefore, effective battery reuse and recycling processes are very important as batteries contain important metals [74-75].

According to the first-order reaction of the lead-acid charge-discharge process, the concentration of SO_4^{2-} ions decreases during discharging as-

$$Pb + PbO_2 + 4H^+ + 2SO_4^{2-} \rightleftharpoons 2PbSO_4 + 2H_2O \quad \ldots (1)$$

This is due to the change in the solubility of lead sulfate in the electrolyte. In the relevant range for lead-acid batteries, the solubility rises with falling electrolyte density (as shown in following Figure 6).

19

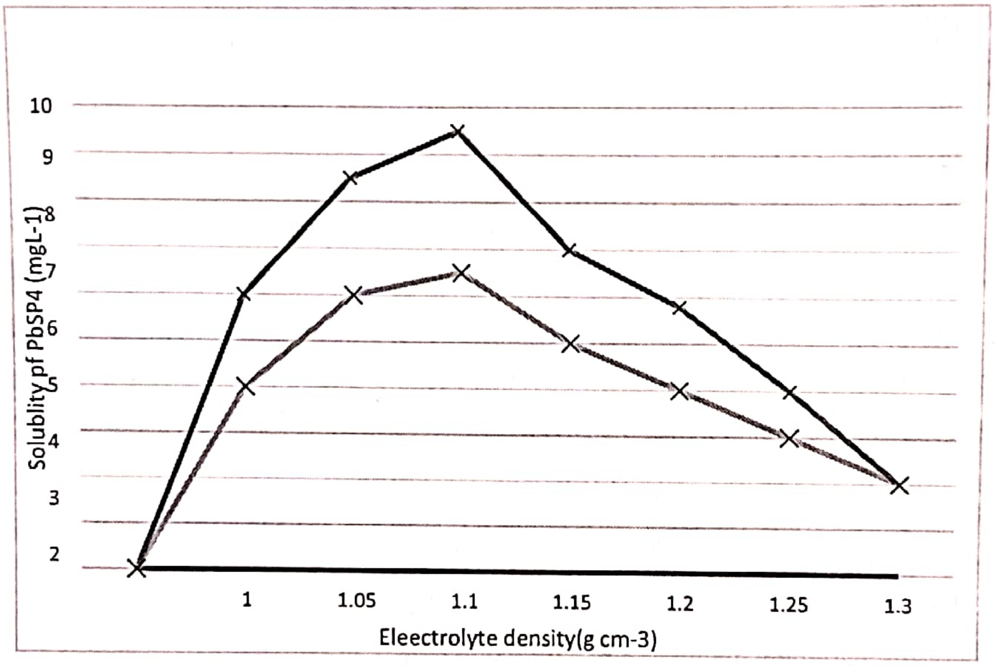

Figure 6: Solubility against electrolyte density

Several ageing processes are accelerated as a result, as opposed to when a battery is fully charged: Sulfation, dendritic formation that could result in mini short circuits, and corrosion during rest times are the first three. As there is increased solubility, large number of sulphate crystals form more quickly. This is brought on by the fact that small crystals have more surface tension and, consequently, surface energy than large crystals.

With this in mind, from an energetic point of view, Pb2+ and SO42− ions are more likely to form on the surface of large crystals than small crystals. Also, lead sulfate molecules are more likely to desorb from small crystals than from large crystals. As the same number of dissolution and crystallization processes occur under equilibrium conditions, a transition of molecules from smaller to larger crystals occurs. Eventually, small crystals will die out, but larger crystals will usually continue to grow. The higher the dissolution rate, the faster the process will go. The term "recrystallization" refers to this process. Sodium sulfate is added to reduce this effect. Dissolving sodium sulphate forms 2Na+ and SO42-, increasing the sulphate content in the electrolyte [76]. Sodium sulfate is used as an electrolyte additive because Na+ ions do not adversely affect the battery. As a result, the sulphate concentration steadily increases, reducing the solubility of the lead sulphate crystals and extending battery life, especially in the case of deep discharge.

20

However, high solubility can also be advantageous. This will quickly remove lead sulfate crystals. This effect was applied decades ago by washing used electrodes with fresh water. Due to the high solubility of lead sulfate in pure water, old and large sulfate crystals could be removed. Theoretically, charging at a temperature of 40-50°C has been shown to be more effective and faster, and promote the dissolution of large amounts of old sulphate crystals. However, increased solubility also occurs at higher temperatures. This is a classic example of the conflicting effects of the same situation and different battery operating modes. Higher solubility is beneficial during charging, but accelerates the formation of large, inert sulfate crystals during resting periods.

The influence on various ageing effects depends on the solubility in and of itself as well as the retention period under the given circumstances.

One of the easiest ways to prevent battery sulfation is to store your batteries properly. When storing batteries, they must be sufficiently charged so that they do not drop below 12.4 volts even when stored fully charged [77]. Application of this maintenance can be achieved by machine learning by providing the data that this charge maintenance prevents sulfate accumulation. Additionally, batteries should not be stored at temperatures above 75°C, as the self-discharge rate doubles for every 10°C increase in room temperature.

1.4.1 Voltage limit

Design parameters optimization in Lithium-ion batteries in experiments like materials selection, cell manufacturing and operation. In lithium-ion batteries, metals like cobalt, nickel, and manganese, which are hazardous and can harm water supplies and ecosystems if they leach out of landfills, are found [78]. Here, our primary aim is to maximize battery lifetime; although, it is a time-consuming process [79-81]. Also, high sample variability requires a large number of experiments. The main challenge is therefore to reduce both the number and time of experiments. To this end, we propose a machine-language method to imaginatively optimize the parameter space that specifies the voltage profile of a fast-charge protocol to maximize battery life [82-83].

Researchers analyzed the accuracy and performance of various SOH (state of health) models to determine the best way to prove the inherent quality of real battery health. Data-driven modeling has important advantages over traditional techniques to improve understanding and prediction of health behaviors. Machine learning technology is an

21

innovative, low-cost, highly accurate and reliable tool for understanding and analyzing battery health.

Much research has focused on applying innovative machine learning (ML) techniques to develop powerful models based on experimental battery data. This work contributes to a variety of things, including whether his SOH model for batteries is robust and validating the behavior of the model.

Researchers have proposed a data-driven alternative model. This model was based on experimental data from a battery and was learned from the data using a machine learning approach. The machine learning techniques in this work were developed by Khalid Akbar et al. Achieved. al. [84] Use Algorithm 1: An algorithmic machine learning engine integrated into this algorithm was connected to a custom script created for this study. Details of the algorithm are given below [84]:

Step 1: Obtain (battery data)

Step 2: Calculate (SOH from charge capacity)

Step 3: while (data) do # check in every dataset

Step 4: divide (data)

Step 5: invoke (Machine Learning algorithm which gives best performance)

Step 6: create (model), verify (model)

Step 7: return (model)

Step 8: display results

Step 9: end of while loop

Step 10: Stop the execution

Algorithm-1 [84]

In this paper, we explored a variety of linear and nonlinear techniques ranging from linear regression to extra trees and found that the boosting method performed best compared to experimental datasets. Furthermore, the best approach for the battery dataset was decision tree regression (DTR) obtained using the boosting method.

The training set and test set must be separated from the dataset in order to execute machine learning, as this study underlines. The computer evaluated its learning performance by applying what it had learnt from the training set to the test set. The test set was fitted to this model to calculate its accuracy and error rates. Finally, a forecast of fresh data points was made using

22

the constructed model. Additionally, the accuracy of data prediction was too high that the error score was around 10^{-2}. This minimal data prediction score verified the output data.

1.5 Study of various machine learning methods used

In this section, we describe the effective theories of the most used Machine Learning methods in Battery management System.

Paper [88] described battery SoC OCV data collection and implemented regression techniques using curve fitting and machine learning concepts. It also covers the basics of battery modeling and the regression techniques used to form the static His SoC-OCV relationship.

Paper titled "Deep Learning Method for Online Capacity Estimation of Lithium-ion Batteries" by Sheng Shen et al. [89] attempted deep learning for online capacity estimation of lithium-ion batteries. They used 10 years of daily cycle data from eight embedded lithium-ion cells and half-year cycle data from 20 18650 lithium-ion cells to validate the performance of their proposed deep learning method. Did. In this paper, compared with traditional machine learning methods such as flat neural networks and relevance vector machines (RVM), their proposed deep learning method has shown higher accuracy and performance in online estimation of lithium-ion battery capacity. It states that it has been shown to provide robustness. Additionally, the next section of this article describes various machine learning methods that can help extend life in battery management systems.

1.6 Methodology

Aim of the research is to analyze and propose a ML (Machine Learning) framework/model for prediction in Various types of battery (power supply resource) domain. The first step in the methodology for Machine Learning research work starts with the preparation of data using Exploratory Data Analysis (EDA). Further step is to clean the data like removing outliers, changing categorical variables and handling imbalanced datasets. For training the model, this research work then uses various machine learning algorithms like supervised and unsupervised. Finally, the model is evaluated using different metrics like recall, f1-score, accuracy, etc.

23

1.6.1 Machine Learning

Machine learning mainly falls into three categories: supervised, unsupervised, and augmented. Supervised learning uses tagged data to train an algorithm. Supervised learning can be divided into two parts: regression and classification. Regression is a form of supervised learning that uses labeled data and uses that data to make predictions in a continuous fashion. Regression problems include types in which the output variables are specified as real numbers.The form of this problem often follows a linear form. Classification in data mining methodology aims at building a model from a training data set. This model can be used to classify records with unknown class specifications. A classification algorithm attempts to determine the class or category of data presented to it. Objects can often belong to multiple categories, and the AI must determine what those categories are and how much confidence thealgorithm has in its predictions.

Unsupervised learning is the second type of machine learning. Here the data is unlabeled and used to train the algorithm. Unsupervised learning can be divided into two parts:clustering and dimensionality reduction. Clustering is a data mining technique that divides a data set into discrete groups. This task divides a population or data points into groups so that data points in the same group are more similar to other data points in the same group than datapoints in other groups That's it. The aim is to separate groups with similar characteristics and assign them to clusters. The k-means clustering algorithm is faster than other clustering

methods for that type. Shorter execution time. Works better with both small and medium data. In our case the dataset is moderate. k-means is the most popular due to the "simple" implementation of the algorithm. Easily identify important relationships between features and extract multivariate or compound variables. No defined number of clusters is required. The optimal number of clusters can be set using an "elbow plot" implemented for a given set of accident data. Dimensionality reduction reduces the dimensionality of the data to remove unnecessary data from the input. Used to remove unnecessary features of data. There are many methods/algorithms for dimensionality reduction in machine learning, among them principal component analysis (PCA), linear discriminant analysis (LDA), kernel principal component analysis (KPCA), and others.

24

Machine learning is a category of algorithms that enable software applications to more accurately predict outcomes without being explicitly programmed [85-86]. The analysis of measuring factors which affect batteries is possible using the Machine Learning model. It is one of the fast-emerging trends which can be helpful for analysis to find out the pattern and to effectively reach the root cause of the lifecycle of batteries (battery issues). Machine Learning is also useful to make predictions to address the measuring factors of a battery so as to enhance its life span. This can be done by acquiring knowledge from previous data by using Machine Learning.

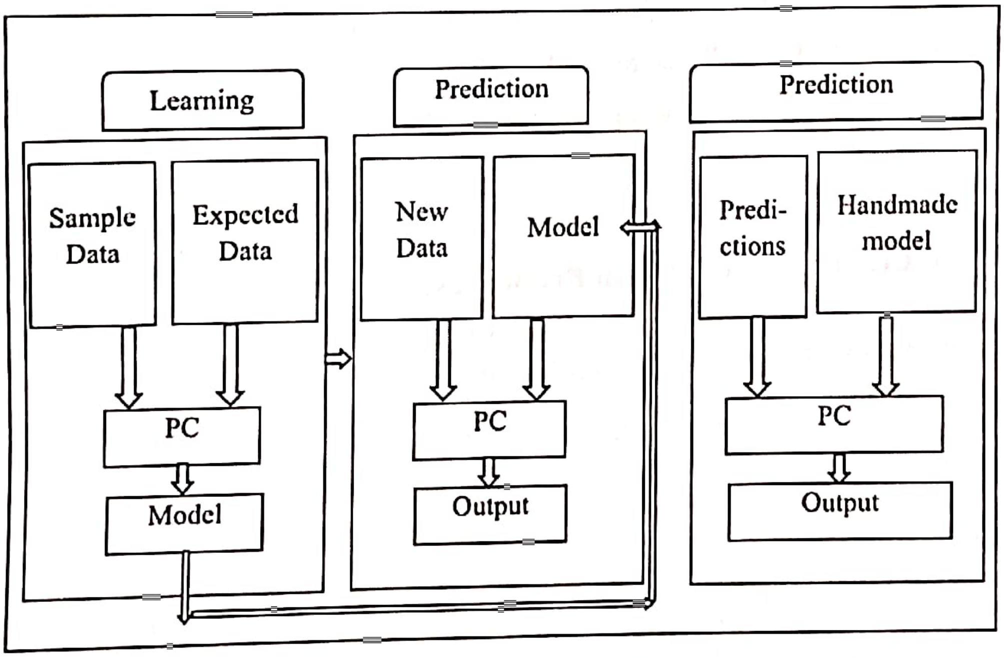

Figure 7: (a) Machine Learning Model Vs. (b) Traditional Model

1.6.2 Machine Learning Model Flow

Figure 7(a) shows how the machine learning model differs from the traditional model shown in Figure 7(b). As shown in the diagram above, the "learning stage" of machine learning modeling, the "model", is one of the inputs to the prediction stage, along with the "new data" to get the output, unlike traditional modeling. We divide our task into two phases: "learning" and "predicting". Training involves preprocessing to clean up the raw data.

1.6.3. Machine Learning Models for Analysis and Predictions

In order to measure the factors that affect batteries with perspective of efficiency and productivity- Machine Learning algorithms are used to identify issues like: sulphation and voltage limits. Hence, this research review proposes the regression model (Machine Learning technique) over the attributes. Typically, the performance of the regression model was

25

evaluated using the root mean square error (RMSE)

The regression model pinpoints measurement-related problems, but it's critical to know which variables affect the battery's lifespan in order to take action to solve these problems. The Association Rule Mining (ARM) algorithm is thus used in this article to give a framework for examining battery productivity and efficiency. The ARM pinpoints the relationships between different battery characteristic metrics that adhere to certain minimal support and confidence requirements. The ARM uses an unsupervised learning approach without the use of class labels. According to Figure 4, both supervised and unsupervised learning include 'classifying tasks' to some extent. The model utilized in this study, built using an unsupervised learning approach, is fine-tuned using labelled data. When research studies use flagged data to refine the models they use, then the algorithm is trained using unsupervised learning – such as the use of different hyperparameters.

1.6.4 Machine Learning Model with Preprocessing

Following Figure 8 shows the flow of Machine Learning Model-

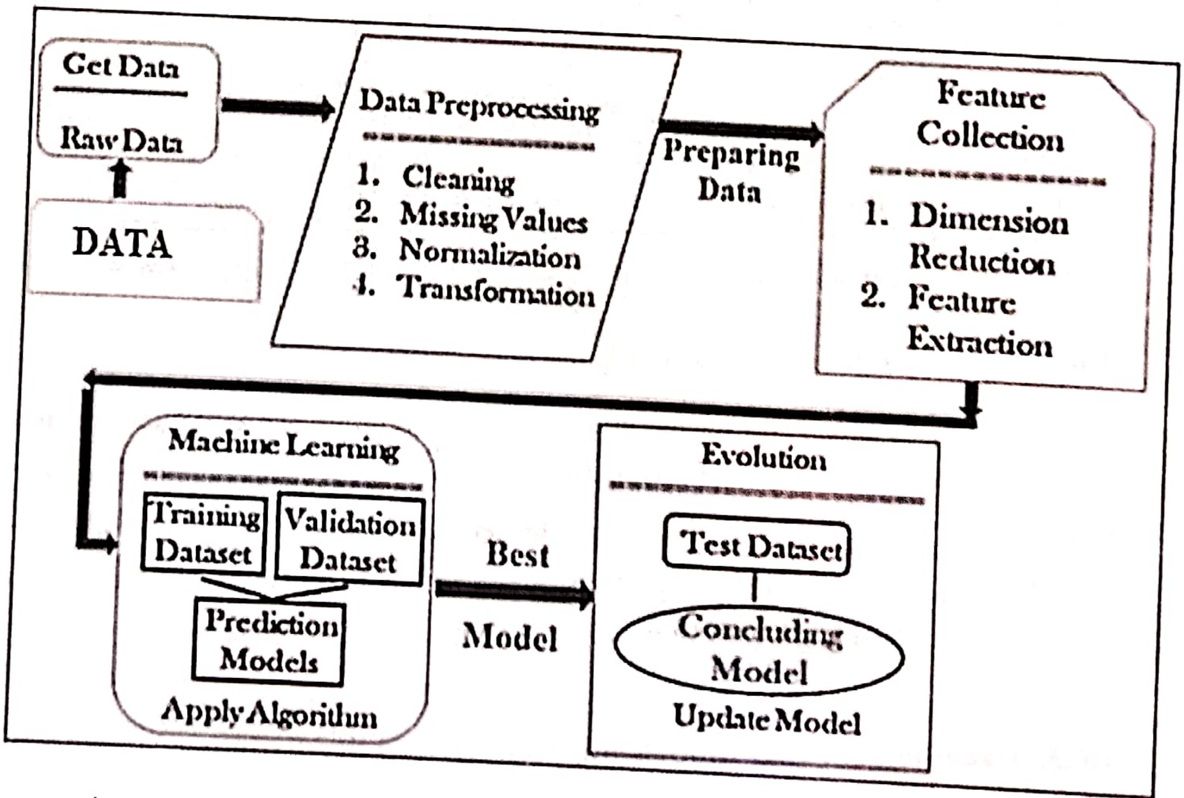

Figure 8: Machine Learning Model flow

As shown in the above Figure 8, Machine Learning model includes pre-processing. It involves formatting, cleaning, and sampling. The data sanitization or data cleaning process

26

identifies and removes inaccurate records. This is where you detect incomplete, unreliable, inaccurate, or irrelevant parts of your data, and recover, refactor, or remove dirty and coarse data. In this work, preprocessing minimizes unnecessary data and yields better processing results. In this way the collected data were preprocessed and filtered. Data normalization is done by standardizing attributes using the following formula:

$$X_{standardization} = \frac{X-mean(X)}{Standard\_deviation(X)} \quad \text{.... (2)}$$

The transformation of raw data into features suitable for modeling is done in the feature extraction part. Data preprocessing and feature extraction have a significant impact on model performance.

Datasets help predict future data outcomes by analyzing models. This is where overfitting by i occurs. presence of noise, ii. a limited training set size, and iii. Classifier complexity. The test set represents the dataset used to test the predictive model produced by the training set to avoid overfitting [87]. Overfitting is likely to occur when the number of features is similar to or greater than the number of observations stored in the dataset. Therefore,we propose a dimensionality reduction technique to avoid this problem. During the feature gathering process, feature extraction can reduce the number of features in the dataset by creating new features from existing features, while feature selection removes unnecessary features. As shown in Figure 8, after feature collection – ML (Machine Learning) algorithms i.e.,Regression, Classification, Clustering, etc. plays an important role in the analysis and the prediction.

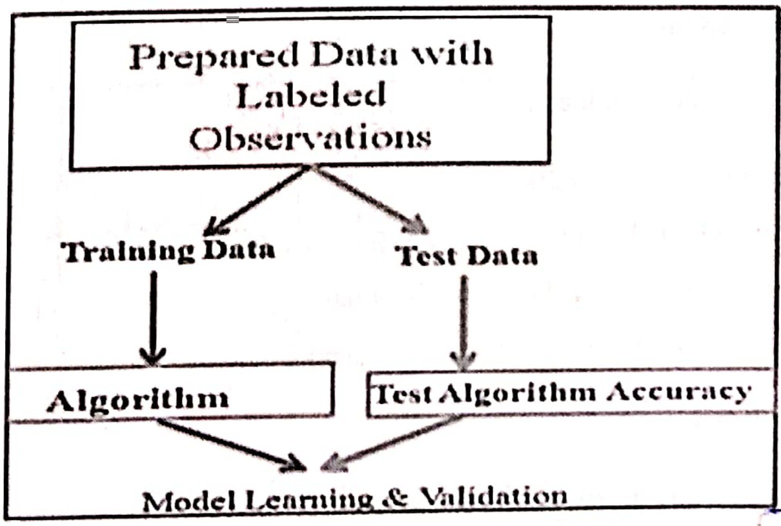

Figure 9: Test-Train Split

RMSE is used to ensure and evaluate the performance of the algorithm used. Test data is usedto see how well the machine can predict new responses based on training. Data tests are performed to obtain the best model.

27

1.7 Working on Measuring factors

1.7.1 Sulfation

A lead sulphate is formed on both negative and positive plates when a cell battery is discharged. This sulfate is converted into active materials once the charging is happening. During the operation of a lead acid battery, such type of sulfation is vital. This lead sulfate is highly conductive that consists of small crystals. Self-discharge results into lead sulfate. In such situation, the active material of the plates is converted into lead sulfate. In such cases, the lead sulfate changes to crusts of lead sulfate. However, such type of sulfation deposition because of the misuse of the battery or ignorance can cause irreversible damage to the battery plates.

This type of irreversible sulfation arises due to: [90]

- a battery is kept in a discharged state for a substantial period of time
- a battery is added with acid instead of water
- a battery is repeatedly over discharged or is continually undercharged
- a battery has low electrolyte in the cells due to insufficient topping up
- a battery is operated at higher temperatures – beyond the advised safe limits of operation
- a battery is not repaired in time as soon as it is observed that one or more of its cells is found to be lagging

The indications of sulfation of the plates are: [90]

- a decrease in battery capacity
- start of gassing much before during the charge cycle
- color of the positive plates becomes abnormal – light brown with white spots
- negative plates volume increases with substantial bulging

1.7.2 Voltage limit

An article by Yang, R. [91] defined a data-driven approach in which SOH can be estimated using cycles of discharge voltage and load current.

The work of Peter M. Attia [92] shows an extended data table using several protocols given below. Of his nine fast-charging protocols tested by researchers, three are the top three

28

CLO-estimated protocols, four are based on approximations to multi-stage fast-charging protocols from the battery literature, and two are CLO. was selected to cover the range of - Estimated cycle life as shown in the table. Four protocols were created by calculating current ratios between different steps and converting these ratios to his 10-min fast-charge range. The protocol was based on an approximation of the multi-step fast-charging method found in the battery literature. The voltage limit corresponded to the charging method.

While retaining similar current ratios as the literature protocols, the researchers modified the voltage restrictions and charging timeframes to meet their methods. For each charging process, five batteries were tested in order to have a good idea of the genuine cycle lifetimes. During testing, the out-of-threshold prediction interval test was removed (channel 12; 3.6C-6.0C-5.6C-4.755C). Extended Data compared three different approaches for calculating cycle life results: CLO, preliminary prediction from validation, and final measurement from validation [92].

Charging protocol	CLO-estimated cyclelife	Early-predicted cycle life (from validation)	Final cycle life (from validation)	Source
3.6C-6.0C-5.6C 4.755C	1103±131	1013±115	755±81	Zhang[93]
4.4C-5.6C-5.2C 4.252C	1174 ±76	1056±127	884 ±132	Protocol with third highest CLO-estimated mean cycle life.

29

4.8C-5.2C-5.2C-4.160C	1185±78	1047± 49	890 ±90	Protocol with highest CLO-estimated mean cycle life.
5.2C-5.2C-4.8C-4.160C	1183±86	1098±134	912±118	Protocol with second-highest CLO-estimated mean cycle life.
6.0C-5.6C-4.4C-3.834C	954 ±164	963± 26	880 ±85	
7.0C-4.8C-4.8C-3.652C	876±183	964 ± 43	870 ±70	Samsung Patents[94, 95]
8.0C-4.4C-4.4C-3.940C	818± 212	854 ± 44	702±51	Nottenet al.[96]
8.0C-6.0C-4.8C-3.000C	775± 273	698± 40	584 ±60	Tesla Patents[97, 98]
8.0C-7.0C-5.2C-2.680C	648±174	580 ±68	496± 49	

Table 1: Extended Data Table [92]

30

As shown in the table above, the columns are the CLO estimated average cycle life for each protocol, the initial predictions in the validation experiment, and the final tested cycle life. For the CLO-estimated cycle life, the error represents the CLO-estimated standard deviation after round 4 (σk,4). Errors between early predicted cycle life and final cycle life from validation represent 95% confidence intervals (n=5; however, n= Four). Two unsourced logs were selected to obtain a representative sample from the distribution of CLO estimated cycle life. The literature on fast-charging protocols is taken from refs [93-98].

1.8 Regression Analysis

Regression analysis is a statistical model used to predict numerical data rather than labels. Regression models return prediction results as continuous values, while classification models predict discrete outputs. You can also identify distribution trends based on available or historical data. Here, the task of regression is to predict the "measured attribute" of a given study area. Despite previous work, it remains difficult to accurately predict battery life from state-of-charge (SOC) and state-of-charge (SOH) estimates under environmental and load conditions that differ from training data. known for Configuration. Advanced regression, classification, and state estimation algorithms play an important role here [99].

There are several variables/attributes that come into play in regression, such as the dependent variable (the main variable you are trying to understand) and the independent variable (factors that can influence the dependent variable). For regression analysis to work on this problem, all relevant data related to the battery must be collected. It can be plotted on a graph with an x-axis and a y-axis.

The main reasons we use regression analysis:

1. To determine the relationship between two or more attributes,
2. To understand how one variable change when another change, and
3. To predict and identify the measuring attributes to enhance the lifespan ofbattery by considering the attributes and dataset.

There are many types of regression analysis. In this research paper we will consider two: linear regression and multiple regression.

31

1.8.1 Linear Regression

Also called "Simple Linear Regression". Establishes a linear relationship between two variables. It also tries to draw the closest line to the data by finding the slope and intercept that define the line and minimizing the regression error. Linear regression presents the task of finding the output 'y'. H. Dependent variable based on given input "x", i.e. H. Independent variable to predict. Hence it is called linear regression. Figure 10 below shows a straight line when the independent variable is plotted on the x-axis and the dependent variable on the y-axis.

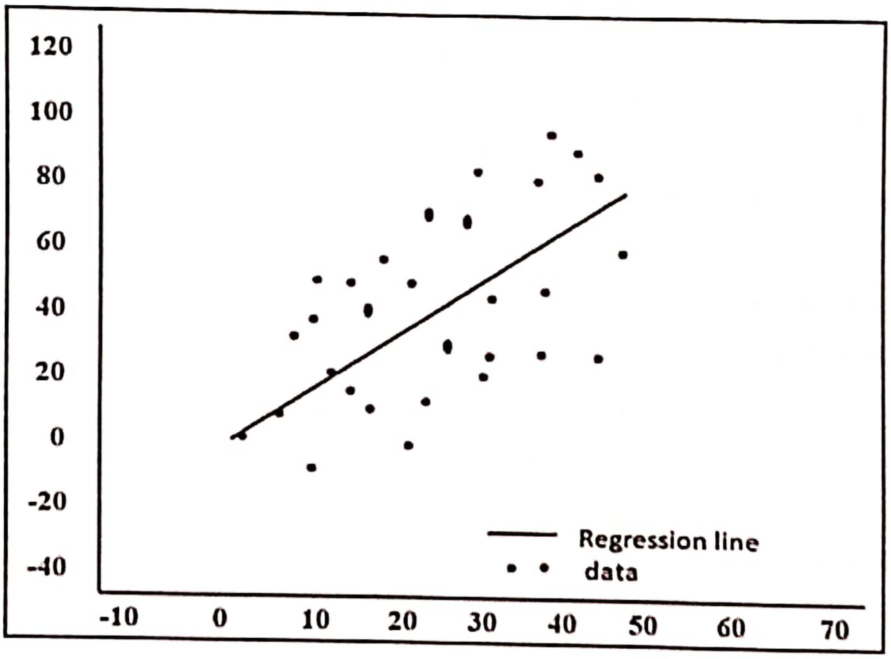

Figure 10: Linear Regression

Source: Python Programming Tutorials

The equation of the above straight line is:

$$y = mx + b \quad ....(3)$$

where $'b'$ is the intercept and $'m'$ is the slope of the line. Thus fundamentally, in two dimensions, the Linear regression algorithm provides us the most optimal value for the intercept and the slope. The variables $'y'$ and $'x'$ are the data features and cannot be changed. The values $'b'$ and $'m'$ that we can control.

Many data relationships do not follow straight lines, so statisticians use nonlinear regression instead. The two are similar in that they graphically trace specific responses from a set of variables. These nonlinear models are more complex than linear models. This is because functions are created by a set of assumptions that can be made by trial and error.

32

1.8.2 Multiple Regression

When two or more explanatory variables are linearly related to the dependent variable, the regression is called "multiple linear regression." This is a broader class of regressions, including linear and nonlinear regressions with multiple explanatory variables. For example, consider a scenario where multiple features need to be predicted.

For example, the dependent variable (target variable) is dependent upon several independent variables. A regression model involving more than two variables as hyperplane. Its equation is as follows:

$$y = b_0 + m_1b_1 + m_2b_2 + m_3b_3 + .. \; m_nb_n \; .... (4)$$

Thus, in two dimensions a linear regression model is a straight line; in three dimensions it is a plane, and in more than three dimensions, a hyperplane.

For performance evaluation of an algorithm, we use following: The mean of the absolute value of the errors is known as Mean Absolute Error (MAE). It is calculated as:

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j| \quad ... (5)$$

The mean of the squared errors is known as Mean Squared Error (MSE) and is calculated as

$$MSE = \frac{1}{N}\sum_{i}^{n} (Y_i - \hat{y}_j)^2 \quad .... (6)$$

The square root of the mean of the squared errors is known as Root Mean Squared Error (RMSE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n} (y_j - \hat{y}_j)^2} \quad ....(7)$$

We then performed a regression analysis of the total number of events divided by the dependent variable using the selected independent variables. Estimate relationships between dependent and independent data using regression models, a supervised learning approach. Here it may be useful to predict the limits of the measurement factors, namely sulfation and stress. It is used to identify the factors that have the greatest impact on the predicted output. It is also used to examine how various variables relate to each other.

Therefore, if you have multiple independent variables and one dependent variable, it is recommended to use "Multiple Linear Regression" or "Multivariate Regression". Multiple regression is an extension of simple linear regression. Used to predict the value of one variable based on the values of two or more other variables. Note here that the variable we want to predict is called the dependent variable.

1.9 Prediction using Clustering

Clustering and classification are two pattern recognition techniques used in machine learning. Both techniques have certain similarities, but the difference is that classification uses a defined class to which objects are assigned, whereas clustering identifies similarities between objects.

We also propose a clustering composed of different sectors to identify population groups. This review article gains prior knowledge of the types of battery attributes required for battery protection and life extension systems by comparing the variables with those of the cluster. In this study, we use the 'elbow method' to find the optimal number of clusters and characterize each cluster. In this situation, the model generation with value 'k' is identified. After obtaining the number of clusters to create, we propose a k-means algorithm to segment the dataset. After obtaining a good segmentation of the dataset, the next task is to perform the characterization of each cluster. A thorough analysis of each cluster reveals a battery variable, that is. The attributes that classified the H. clusters were sulfation, stress limit, corrosion, passivation, and thermal runaway.

Hence, the paper carried out analysis, visualizations, and classification of the incidents using k-means clustering. A detailed study of each cluster discovers composite variables.

Clustering is the task of dividing a dataset into groups called clusters [100]. That is, objects are grouped according to these common characteristics and distinguished from other objects called clusters. The purpose of clustering is to split the data so that points in one cluster are very similar and points in different clusters are different. Usually you decide to group under unlabeled data.

Using a clustering algorithm means giving the algorithm a large amount of input data and letting it find all the groups in the detectable data, these groups are called clusters. A cluster is a group of data points that are similar to each other due to their relationship to

34

surrounding data points. Clustering is used for things like feature engineering and pattern discovery.

Clustering may be an excellent place to start when you are starting with data that you are unfamiliar with. Companies that wish to apply client segmentation, define groupings, and concentrate on certain products or services often need to collect frequent features from their customers. Thus, a firm can defend a specific campaign, service, or product if a sizable portion of its clients have certain characteristics (gender, age, etc.). Clustering is useful for locating comprehensive data and insights in this way. Partitional and hierarchical clustering are the two clustering methods that have received the greatest attention and are utilized the most. These algorithms have been largely used in an extensive range of applications mainly due to their simplicity and ease of implementation relative to other clustering algorithms [101].

The purpose of partition clustering algorithms is to discover groups present in the data by optimizing a specific objective function and iteratively improving the quality of the partitions. These algorithms are also called prototype-based clustering algorithms because they mostly require specific user parameters to select the prototype points that represent each cluster.

In contrast, hierarchical clustering algorithms tackle the clustering problem by developing a binary tree-based data structure. It's called a dendrogram. Once the dendrogram is built, the appropriate number of clusters can be automatically obtained by splitting the tree at different levels to obtain different clustering solutions for the same dataset without having to run the clustering algorithm again. can be selected to Hierarchical clustering can be achieved by his two different methods: bottom-up clustering and top-down clustering.

As shown in Figure 11, the hierarchy is further divided into two subdivisions: split and agglomerated. Division algorithms are different from aggregation algorithms. Splitting starts with all the points in the cluster and splits them to create more clusters. These algorithms build a distance matrix of all existing clusters and perform inter-cluster linking according to link criteria. The clustering of data points is represented by a dendrogram [102].

Hierarchical and partitioned approaches use the concept of a dendrogram, but can return significantly different result sets depending on the criteria used in this clustering process. A splitting method requires a set of initial seeds (or clusters) to be specified. These are iteratively improved. As shown in Figure 11, split clustering is further divided into his two parts: k-means clustering and fuzzy c-means clustering.

35

in the same cluster are as similar to each other as possible. The approach is maximizing expectations for solving problems.

k-means works as – group the items into 'k' clusters such that all items in the same cluster are as similar to each other. The algorithm [104] is as follows:

1. Within the data domain, k initial 'means' are randomly generated,
2. k-clusters are formed by correlating each observation with the nearest mean,
3. The centroid of each of the k-clusters turn out to be a new mean,
4. Steps 2 and 3 are repeated until convergence has been reached.

1.9.1. RUL based k-means clustering

Thousands of batteries are in use around the world, and their lifespan degrades rapidly over time. Therefore, it is very important to check the state of deterioration of the battery before the first failure occurs during the discharge cycle, and it is a very difficult task even in practical use. Therefore, clustered predictors are proposed to improve the accuracy of battery state prediction. Haider et al. In the paper, his 1-year historical data of his 40 batteries in a large data center are presented to verify the effectiveness of the proposed methodology [105].

A clustering prediction framework based on the k-means algorithm has been proposed to improve the accuracy of predicting the RUL of lithium-ion batteries. The data are differentiated by his two cross-cycle health factor analysis using the k-means algorithm [106].

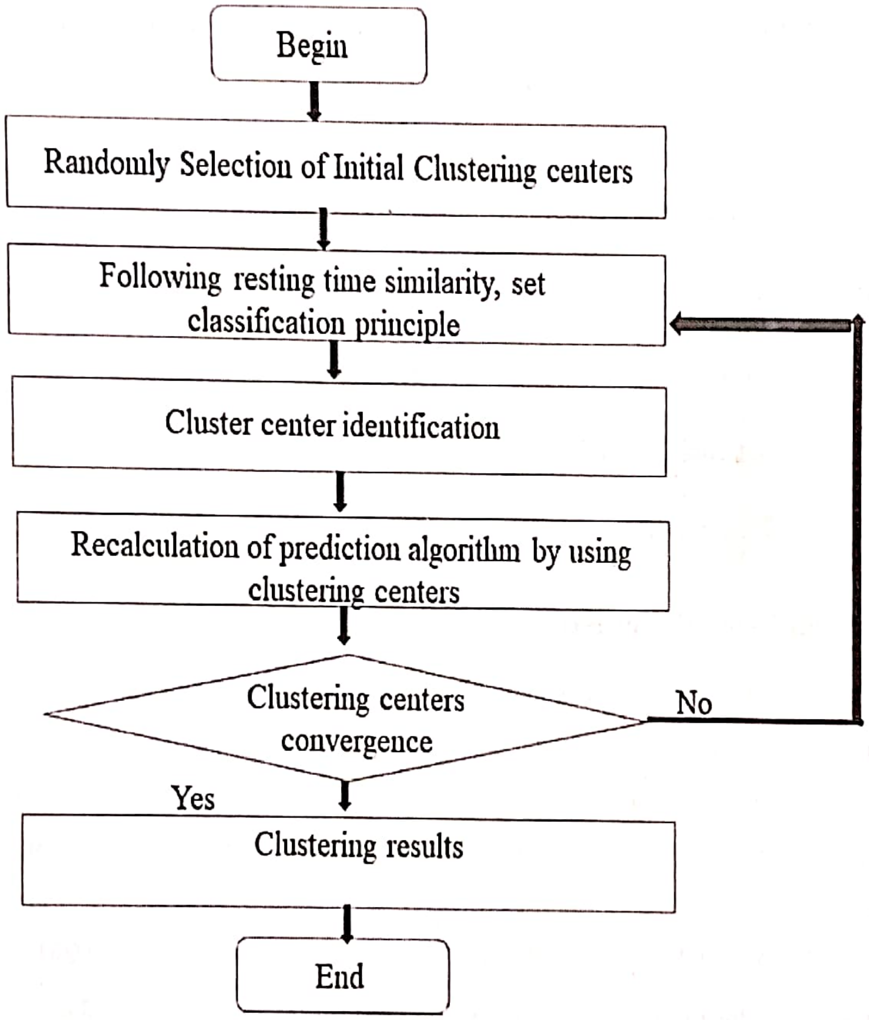

Figure 12: Battery life model based on a Clustering framework [106]

The degradation mechanism of lithium-ion batteries is complex. Therefore, as the model accuracy increases, more complex conditions are added. Different batteries have different material compositions, which reduces the robustness of the model. As a result, many studies on RUL prediction have been done from the perspective of data-driven methods. This RUL is a purely subjective estimate of the number of years a battery is said to have served its intended purpose before it needs to be replaced.

Mathematical modeling of batteries is important as charging/discharging techniques improve and battery capacity increases [109]. Data-driven techniques apply statistical or machine learning to large amounts of data. Rather than building complex physical models, these methods look for functions of battery charge/discharge cycles and electrical measurements of RUL [110]. These models extrapolate predictions from similar learnings

38

about battery capacity depletion trends or predict system states based on previously observed data. His two main types of data-driven RUL prediction approaches are feature-based her RUL prediction and autoregressive capacity fading prediction [111]. Based on the captured capacity fade curve, the autoregressive prediction of capacity fade is used to estimate the cycle life of the battery. These studies used previously developed capacity fading state methods such as SVR [112], filtering algorithms [113–114], long short-term memory networks [115], and Gaussian process regression [116] to perform capacity prediction. do. Several studies [117-118] have attempted to use neural networks or regression methods to determine remaining battery life from the intersection of voltage, current, and temperature. With further developments in algorithms and neural networks, more advanced forecasting methods are now being used in practice [119]. Several combined methods have been proposed to predict RUL. The algorithm achieves excellent prediction accuracy by maximizing the use of many models. Various forecasting techniques have been combined with empirical model decomposition approaches [120-121]. Support vector classification and regression attributes were integrated to estimate the RUL of Li-ion batteries [117]. These techniques emphasize the advantages of algorithms by considering combinatorial algorithms. Many scientists recommend using clustering techniques. Clustering is currently used in various forecasting areas, such as forecasting wind speed [124], bus load [123], and energy consumption [122]. Clustering has not been used to predict the remaining lifetime of lithium-ion batteries due to lack of sufficient evidence to measure capacity fade in lithium-ion batteries.

1.9.2 BMS using clustering

As the charge and discharge times lengthen, Li-ion Batteries become less effective or even fail. The gadget won't operate normally if adequate action is not taken before a battery fails, which could seriously compromise safety [107-108]. Therefore, these issues may be mitigated by a reliable battery life prediction system. It is challenging for humans to create a solid rule-based algorithm to examine capacity fade since degradation in Li-ion Batteries is not directly reflected in it [125].

Battery management systems are designed based on characterization criteria to check battery health and provide a level of safety to ensure predictable functioning of lithium-ion batteries. Recently, more researchers have started to look at the prospects for the remaining useful life of lithium-ion batteries [126]

1.10 Conclusions and Perspective

This work reviewed some of the key issues related to measuring factors for overall BMS. On the very onset, important findings are exhibited in the estimation of SOH and SOC using the Machine Learning method.

Battery behavior is non-linear and cannot be predicted directly, so an indirect approach of estimation and prediction is required. A number of features were observed from the discharge curves of different batteries at different current and temperature levels to estimate her SOH and SOC of the battery. For example, battery B0006 has a current and temperature of 2.0A and 297.15K, while battery B0047 has a current and temperature of 1.0A and 277.15K. Battery arrays mentioned in various studies were taken from other data sources. Some studies used data from battery 40, while others used data up to battery 27 to build the model. These models were created to understand the behavior of battery systems. Others have developed compact models to understand the combined behavior within a system. The machine learning model takes the estimated properties and outputs SOH over all battery discharge cycles in terms of pseudolinear range and arm length. Also, the area under the voltage trace is useful when estimating SOC%. In addition to estimating accurate SOC% for depletion cycles, this method can produce a downward trend for all training cycles. That is, I propose to use the voltage of the battery as a reference point and use a reverse engineer to obtain the SOH. Given a battery at a particular point in time, estimate of SOC is possible. For large datasets, the error obtained is negligible.

Despite the fact that polynomial regression models provide the best RMSE compared to other machine learning approaches, this may not be suitable for large datasets. As a result, improved techniques such as ANNs and clustering may prove useful in the future. Finally, ANN is better suited for online SOH estimation because the proposed model requires the entire discharge curve to calculate his SOH and SOC of the battery using a larger dataset.

References

[1] B. Dunn, H. Kamath, J. Tarascon, Electrical Energy Storage for the Grid: A Battery of Choices, Science 2011, 334, pp. 928–935.

[2] X. Hu, C. Zou, C. Zhang, Y. Li, Technological Developments in Batteries: A Survey of Principal Roles, Types, and Management Needs, IEEE Power Energy Mag 2017, 15, pp. 20–31.

[3] X. Han, M. Ouyang, L. Lu, J. Li, A Comparative Study of Commercial Lithium-Ion Battery Cycle Life in Electric Vehicle: Capacity Loss Estimation. J. Power Sources 2014, 268, 658–669.

[4] H. Tian, P. Qin, K. Li, Z. Zhao, A review of the state of health for lithium-ion batteries: Research

40

status and suggestions, Journal of Cleaner Production, Volume 261, 10 July 2020, 120813, https://doi.org/10.1016/j.jclepro.2020.120813.

[5] H.Shareef, Islam, M.M. Mohamed, A. A Review of the Stage-of-the-Art Charging Technologies, Placement Methodologies, and Impacts of Electric Vehicles. Renew. Sustain. Energy Rev. 2016, 64, 403–420.

[6] Electric vehicle industry in India: Current state, government policies, https://timesofindia.indiatimes.com , [accessed 18 August 2022].

[7] Y. Kim, C. Hwang, E. Kim, C. Cho, State of Charge-Based Active Power Sharing Method in a Standalone Microgrid with High Penetration Level of Renewable Energy Sources. Energies 2016, 9, 480.

[8] H. Ren, Y. Zhao, S. Chen, T. Wang, Design and implementation of a battery management system with active charge balance based on the SOC and SOH online estimation. Energy 2019, 166, 908–917.

[9] X. Tang, Y. Wang, C. Zou, K. Yao, Y. Xia, F .Gao. A novel framework for Lithium-ion battery modeling considering uncertainties of temperature and aging. Energy Convers. Manag. 2019, 180, 162–170.

[10] Z. Wei, F. Leng, Z. He, Zhang, W.; Li, K. Online State of Charge and State of Health Estimation for a Lithium-Ion Battery Based on a Data–Model Fusion Method. Energies 2018, 11, 1810.

[11] M. Zhang, X. Fan, Review on the State of Charge Estimation Methods for Electric Vehicle Battery. World Electr.Veh. J. 2020, 11, 23.

[12] State of Charge determination – Battery and Energy Technologies, Electropaedia, https://www.mpoweruk.com/

[13] C. Li, F. Xiao, Y. Fan, An Approach to State of Charge Estimation of Lithium-Ion Batteries Based on Recurrent Neural Networks with Gated Recurrent Unit, Energies 2019, 5, 1592.

[14] J. Lee, O. Nam, B.H. Cho, Li-ion battery SOC estimation method based on the reduced order extended Kalman filtering. J. Power Sources 2007, 174, 9–15.

[15] M. Talha, F. Asghar, S.H Kim. A Neural Network-Based Robust Online SOC and SOH Estimation for Sealed Lead–Acid Batteries in Renewable Systems. Arab. J. Sci. Eng. 2018, 44, 1869–1881.

[16] J.C.A. Anton, P.J.G Nieto, C.B. Viejo, J.A.V Vilán. Support Vector Machines Used to Estimate the Battery State of Charge. IEEE Trans. Power Electron. 2013, 28, 5919–5926.

[17] S.C. Huang, K.H. Tseng, J.W. Liang, C.L. Chang, M.G. Pecht, An Online SOC and SOH Estimation Model for Lithium-Ion Batteries. Energies 2017, 10, 512.

[18] A. Patil, V. Patil, J. Choi, J. Kim, S. Yoon, January-2017, Solid Electrolytes for Rechargeable Thin Film Lithium Batteries: A Review', Journal of Nanoscience and Nanotechnology 17(1): 29-71

[19] S. Sheikh, M. Anjum, M. Khan, S. Hassan, H. Khalid, A. Ben-Brahim, A Battery Health

41

Monitoring Method Using Machine Learning: A Data-Driven Approach, energies, MDPI, pp. 1-15

[20] S. Chitnis, P. Gokhale, Jan-2020, An approach to analyze and predict highway accident scenarios in India, Test Engineering and Management, ISSN: 0193-4120, Published by: The Mattingley Publishing Co. Inc., pp. 17609 – 17617.

[21] D. Sharma, N. Kumar, 2017, October. A Review on Machine Learning Algorithms, Tasks and Applications, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 6, Issue 10, pp. 1548-1552

[22] C. Park, C. Took, J. Seong ,2018, February. Machine Learning in biomedical engineering, Biomedical Engineering Letters, Springer Link.

[23] M. Brady, L. A. Gerhardt, H. F. Davidson, Eds.; Robotics and Artificial Intelligence; Springer-Verlag New York Inc, 2012.

[24] C. Chen, A. Seff, A. Kornhauser, J. Xiao, DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. Proc. IEEE Int. Conf. Comput. Vis. 2015, 2722–2730.

[25] M. Q. Raza, A. Khosravi. A Review on Artificial Intelligence Based Load Demand Forecasting Techniques for Smart Grid and Buildings. Renewable Sustainable Energy Rev. 2015, 50, 1352–1372.

[26] A. Mistry, A. A. Franco, S. J. Cooper, S. A. Roberts, V. Viswanathan. How Machine Learning Will Revolutionize Electro-chemical Sciences. ACS Energy Lett. 2021, 1422–1431.

[27] L. Kong, M. Pecht, August-2020, A Look Inside Your Battery: Watching the Dendrites Grow, Battery Power

[28] V.S. Dave, K. Dutta, Neural network-based models for software effort estimation: a review, Artificial Intelligence Rev., 42(2),2014, pp.295-307

[29] MA Hannan , MSH Lipu , A Hussain , A Mohamed. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: challenges and recommendations. Renew Sustain Energy Rev Oct. 2017;78: 834–54. https://doi.org/10.1016/J.RSER.2017.05.001.

[30] R Xiong, J Cao, Q Yu, H He, F Sun. Critical Review on the Battery State of Charge Estimation Methods for Electric Vehicles. IEEE Access 2018;6:1832–43. https://doi.org/10.1109/ACCESS.2017.2780258.

[31] DNT How, MA Hannan, MS Lipu Hossain, PJ Ker. State of Charge Estimation for Lithium-Ion Batteries Using Model-Based and Data-Driven Methods: a Review. IEEE Access 2019;7:136116–36. https://doi.org/10.1109/ACCESS.2019.2942213.

[32] Y. Liu, Y. He, H Bian, W. Guo, X. Zhang. A review of lithium-ion battery state of charge estimation based on deep learning: directions for improvement and future trends. J Energy Storage 2022;52:104664. https://doi.org/10.1016/j. est.2022.104664.

[33] E. Galiounas , G. T. Tranter, E. R. Owen, B. J. Robinson, R. P. Shearing, J.L. Dan Brett, Battery

42

state-of-charge estimation using machine learning analysis of ultrasonic signatures , Energy and AI 10 2022, Energy and AI, journal homepage: www.sciencedirect.com/journal/energy-and-ai

[34] W. He, et al., State of charge estimation for li-ion batteries using neural network modeling and unscented kalman filter-based error cancellation, International Journal of Electrical Power & Energy Systems, vol. 62, pp. 783-791, 2014.

[35] M. Charkhgard and M. Farrokhi, State-of-charge estimation for lithium-ion batteries using neural networks and ekf, IEEE transactions on industrial electronics, vol/issue: 57(12), pp. 4178-4187, 2010.

[36]G. Alexander, P. Joris, E. Loubna, Active electrode material, https://patents.google.com/patent/GB2598432A/en

[37] A. Vabalas, E. Gowen, E. Poliakoff, A. J. Casson, Machine learning algorithm validation with a limited sample size, November-2019, https://doi.org/10.1371/journal.pone.0224365

[38] D. Fourches, E. Muratov, A. Tropsha, Trust, but Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. J. Chem. Inf. Model. 2010, 50, 1189– 1204.

[39] D. Fourches, E. Muratov, A. Tropsha, Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. J. Chem. Inf. Model. 2016, 56, 1243−1252

[40] F. Jose Rodrigues Jr, Larisa Florea, C. F. Maria de Oliveira, D. Diamond, O. N. Oliveira Jr, Big data and machine learning for materials science, Discover Materials volume 1, Article number: 12, 2021

[41] M. Nakayama, K. Kanamori, K. Nakano, R. Jalem, I. Takeuchi, H. Yamasaki, Data-Driven Materials Exploration for Li-Ion Conductive Ceramics by Exhaustive and Informatics-Aided Computa- tions. Chem. Rec. 2019, 19, 771−778

[42] A. Bhowmik, I. E. Castelli, J. M. Garcia-Lastra, P. B. Jørgensen, O. Winther, T. Vegge, A Perspective on Inverse Design of Battery Interphases Using Multi-Scale Modelling, Experiments and Generative Deep Learning. Energy Storage Mater. 2019, 21, 446−456.

[43] B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. Science (Washington, DC, U. S.) 2018, 361, 360−365.

[44] A. Bhowmik, I.E.Castelli, J. Garcia-Lastra, P.B.Jorgensen, O.Winther, T.Vegge, A perspective on inverse design of battery interphases using multi-scale modelling, experiments and generative deep learning, Energy Storage Materials, Volume 21, September-2019, pp. 446-456.

[45] N. Kireeva, Pervov, V. S. Materials Informatics Screening of Li- Rich Layered Oxide Cathode Materials with Enhanced Characteristics Using Synthesis Data. Batter. Supercaps 2020, 3, 427−438

[46]P.M. Attia, A. Grover, N. Jin, K.A. Severson, T.M. Markov, Y.H. Liao, M.H. Chen, B. Cheong, N . Perkins, Z. Yang, P.K. Herring, M. Aykol, S.J. Harris, R.D. Braatz, S. Ermon, W.C. Chueh, Closed-loop optimization of fast-charging protocols for batteries with machine learning, Nature, 7795, 2020,

pp. 397-402.

[47]A.D. Sendek, E.D. Cubuk, E.R. Antoniuk, G. Cheon, Y. Cui, E.J. Reed, Machine learning-assisted discovery of solid Li-ion conducting materials, Chem. Mater., 2 , 2019, pp. 342-352.

[48] X. Wang, R. Xiao, H. Li, L. Chen. Quantitative Structure- Property Relationship Study of Cathode Volume Changes in Lithium Ion Batteries Using Ab-Initio and Partial Least Squares Analysis. J. Mater. 2017, 3, 178−183.

[49] R. P. Joshi, J. Eickholt, L. Li, M. Fornari, V. Barone, J. E. Peralta, Machine Learning the Voltage of Electrode Materials in Metal-Ion Batteries. ACS Appl. Mater. Interfaces 2019, 11, 18494−18503

[50] H. Wang, S. Chen, C. Fu, Y. Ding, G. Liu, Y. Cao, Z. Chen, June-2021, Recent Advances in Conversion-Type Electrode Materials for Post Lithium-Ion Batteries, ACS Materials Lett. 2021, 3, 7, 956–977, https://doi.org/10.1021/acsmaterialslett.1c00043.

[51] O. Allam, B. W. Cho, K. C. Kim, S. S. Jang, Application of DFT-Based Machine Learning for Developing Molecular Electrode Materials in Li-Ion Batteries. RSC Adv. 2018, 8, 39414−39420

[52] S. Ul-Hassan, J. Ahamed, K. Ahmad, Analytics of machine learning-based algorithms for text classification, Sustainable Operations and Computers, Volume 3, 2022, pp. 238-248

[53] M. Attarian Shandiz, R. Gauvin, Application of Machine Learning Methods for the Prediction of Crystal System of Cathode Materials in Lithium-Ion Batteries. Comput. Mater. Sci. 2016, 117, 270−278

[54] V. Patil, A. Patil, S. Yoon, J.B. Choi, May-2013, Structural and Electrical properties of NASICON Type Solid Electrolyte Nano-scaled Glass-Ceramic Powder by Mechanical Milling for Thin Film Batteries, Journal of Nanoscience and Nanotechnology 13(5): 3665-8

[55] R. Jalem, T. Aoyama, M. Nakayama, M. Nogami, Multivariate Method-Assisted Ab Initio Study of Olivine-Type LiMXO 4 (MainGroup M 2+-X 5+ and M 3+-X 4+) Compositions as Potential Solid Electrolytes. Chem. Mater. 2012, 24, 1357−1364.

[56] R. Jalem, M.Kimura, M. Nakayama, T. Kasuga, Informatics-Aided Density Functional Theory Study on the Li Ion Transport of Tavorite-Type LiMTO4F (M3+-T5+, M2+-T6+). J. Chem. Inf. Model.2015, 55, 1158−1168.

[57] N. A. Katcho, J. Carrete, M. Reynaud, G. Rousse, M. Casas-Cabanas, N. Mingo, J. Rodríguez-Carvajal, J. Carrasco. An Investigation of the Structural Properties of Li and Na Fast IonConductors Using High-Throughput Bond-Valence Calculations and Machine Learning. J. Appl. Crystallogr. 2019, 52, 148−157.

[58] K. Fujimura, A. Seko, Y. Koyama, A. Kuwabara, I. Kishida, K. Shitara, C. A. J. Fisher, H. Moriwake, I. Tanaka . Accelerated Materials Design of Lithium Superionic Conductors Based on First-Principles Calculations and Machine Learning Algorithms. Adv. EnergyMater. 2013, 3, 980−985.

44

[59] L. Kong, M. Pecht, August-2020, A Look Inside Your Battery: Watching the Dendrites Grow, Battery Power

[60] Z. Ahmad, T. Xie, C. Maheshwari, J. C. Grossman, V. Viswanathan, Machine Learning Enabled Computational Screening of Inorganic Solid Electrolytes for Suppression of Dendrite Formationin Lithium Metal Anodes. ACS Cent. Sci. 2018, 4, 996−1006.

[61] T. Nakayama, Y. Igarashi, K. Sodeyama, M. Okada, Material Search for Li-Ion Battery Electrolytes through an Exhaustive Searchwith a Gaussian Process. Chem. Phys. Lett. 2019, 731, 136622

[62] K. Sodeyama, Y. Igarashi, T. Nakayama, Y. Tateyama, M. Okada. Liquid Electrolyte Informatics Using an Exhaustive Search with Linear Regression. Phys. Chem. Chem. Phys. 2018, 20, 22585−22591.

[63] T. Nakayama, Y. Igarashi, K. Sodeyama, M. Okada, Material Search for Li-Ion Battery Electrolytes through an Exhaustive Search with a Gaussian Process. Chem. Phys. Lett. 2019, 731, 136622.

[64] E. Heid, M. Fleck, P. Chatterjee, C. Schröder, A.D. Mackerell, Toward Prediction of Electrostatic Parameters for Force Fields That Explicitly Treat Electronic Polarization. J. Chem. Theory Comput. 2019,15, 2460−2469.

[65] D. Bedrov, J. P. Piquemal, O. Borodin, A. D. MacKerell, B. Roux, C. Schröder, Molecular Dynamics Simulations of IonicLiquids and Electrolytes Using Polarizable Force Fields. Chem. Rev.2019, 119, 7940−7995.

[66] F. Mo, B. Guo, W. Ling, J. Wei, L. Chen, S. Yu, G. Liang, June-2022, Recent Progress and Challenges of Flexible Zn-Based Batteries with Polymer Electrolyte, batteries-MDPI, pp. 1-17, https://doi.org/10.3390/batteries8060059

[67] M. Xu, T. Zhu, J. Z. H. Zhang. Molecular Dynamics Simulation of Zinc Ion in Water with an Ab Initio Based Neural Network Potential. J. Phys. Chem. A 2019, 123, 6587−6595.

[68] L. D. Ellis, S. Buteau, S. G. Hames, L. M. Thompson, D.S. Hall, J. R Dahn. A New Method for Determining the Concentration of Electrolyte Components in Lithium-Ion Cells, Using Fourier Transform Infrared Spectroscopy and Machine Learning. J. Electrochem. Soc., 2018, 165, A256−A262.

[69] H. Lu, X. Hu, B. Cao, W. Chai, F. Yan. Prediction of Liquidus Temperature for Complex Electrolyte Systems Na3AlF6-AlF3-CaF2-MgF2-Al2O3-KF-LiF Based on the Machine Learning Methods.Chemom. Intell. Lab. Syst. 2019, 189, 110−120.

[70] F. Häse, L. M. Roch, A. Aspuru-Guzik, Next-Generation Experimentation with Self-Driving Laboratories. Trends Chem. 2019, 1,282−291.

[71] Sulfation – an overview, ScienceDirect

45

[72] Crown Battery, What is a Sulfated Battery and How to Prevent It, 2017

[73] Mitigation of sulfation in lead acid battery towards life time extension using ultra capacitor in hybrid electric vehicle, Journal of Energy Storage, Volume 34, February 2021, 102219

[74] Zhang, Y. Wang, B. Cheng, Q. Li, X. Li, Z., Removal of toxic heavy metal ions (Pb, Cr, Cu, Ni, Zn, Co, Hg, and Cd) from waste batteries or lithium cells using nanosized metal oxides: A review. J. Nanosci. Nanotechnol. 2020, 20, 7231–7254.

[75] S.T. Palisoc, E.J.F. Cansino, I.M.O. Dy, C.F.A. Razal, Reyes, K.C.N. Racines, L.R. Natividad, M.T., Electrochemical determination of tannic acid using graphite electrodes sourced from waste zinc-carbon batteries, Sens. Bio-Sens. Res. 2020, 28, 100326.

[76] D.U. Sauer, Secondary Batteries – Lead – Acid Systems, Encyclopedia of Electrochemical Power Sources - Elsevier, Lifetime Determining Processes, 2009, pp.805-815, https://doi.org/10.1016/B978-044452745-5.00137-4

[77] Crown Battery, June-2022, What is Sulfated Battery and How to Prevent it

[78] Environmental impacts of lithium-ion batteries, From Wikipedia, the free encyclopedia, Retrieved from: https://en.wikipedia.org/wiki/Environmental_impacts_of_lithium-ion_batteries.

[79] T. Baumhöfer, M. Brühl, S. Rothgang, D. U. Sauer . Production caused variation in capacity aging trend and correlation to initial cell performance, J. Power Sources 247, 332–338 ,2014.

[80] P. Keil, A. Jossen, Charging protocols for lithium-ion batteries and their impact on cycle life—an experimental study with different 18650 high-power cells, J. Energy Storage 6, 125–141 ,2016.

[81] K. A. Severson, et al. Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy 4, 383–391 ,2019.

[82] S. Ahmed, et al. Enabling fast charging—a battery technology gap assessment. J. Power Sources 367, 250–262 ,2017.

[83] Y. Liu, Y. Zhu, Y. Cui, Challenges and opportunities towards fast-charging battery materials. Nat. Energy 4, 540–550 ,2019.

[84] K. Akbar, Y. Zou, Q. Awais, M. Jabbar A. Baig, M. Jamil, A Machine Learning-Based Robust State of Health (SOH) Prediction Model for Electric Vehicle Batteries, *Electronics* 2022, *11*(8), 1216; https://doi.org/10.3390/electronics11081216.

[85] D. Sharma, N. Kumar, A Review on Machine Learning Algorithms, Tasks and Applications, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 6, Issue 10, October 2017, pp. 1548-1552

[86] C. Park, CC Took, J. Seong, Machine Learning in biomedical engineering, Biomedical Engineering Letters, Springer Link, February 2018.

[87] X. Ying, (2018). An Overview of Overfitting and its Solutions, CISAT, IOP Journal of Physics: Conf. Series, 1168 ,2019.

46

[88] S.Siva, Suriya Narayanan, S.Thangavel, Machine learning-based model development for battery state of charge–open circuit voltage relationship using regression techniques, Journal of Energy Storage, Elsevier

[89] S. Shen, M. Sadoughi, X. Chen, M. Hong, C. Hu, 2019, A Deep Learning Method for Online Capacity Estimation of Lithium-Ion Batteries, Elsevier

[90] G. Papazov, SECONDARY BATTERIES – LEAD– ACID SYSTEMS- Negative Electrode, in Encyclopedia of Electrochemical Power Sources, 2009, Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Pages 576-589.

[91] Z. Ma, R. Yang, Z. Wang. A novel data-model fusion state-of-health estimation approach for lithium-ion batteries. Appl. Energy 2019, 237, pp. 836–847.

[92] P. M. Attia, A. Grover, N. Jin, K.A. Severson, T. M. Markov, Y. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K . Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon and W. C. Chueh, Closed-loop optimization of fast-charging protocols for batteries with machine learning, 2020, Journal: Nature 578 (7795), doi: 10.1038/s41586-020-1994-5.

[93] S. S. Zhang, The effect of the charging protocol on the cycle life of a Li-ion battery. J. Power Sources 161, 1385–1391 ,2006.

[94] Kim, J. M. et al. Battery charging method and battery pack using the same. US Patent Application US20160226270A1 ,2016.

[95] M.-S. Lee, S.-B. Song, J.-S. Jung, & Golovanov, D. Battery charging method and battery pack using the same. US Patent US9917458B2 ,2018.

[96] P. H. L. Notten, J. H. G. Op het Veld, & J. R. G. van Beek, Boostcharging Li-ion batteries: a challenging new charging concept. J. Power Sources 145, 89–94 ,2005.

[97] A. Paryani, Low temperature charging of Li-ion cells, US Patent US8552693B2 ,2013.

[98] V. H. Mehta, & J. B. Straubel. Fast charging with negative ramped current profile. US Patent US8643342B2 ,2014.

[99] K. Goebel, B. Saha, A. Saxena, J. R. Celaya, J. P. Christophersen, Prognostics in Battery Health Management, 2-11, IEEE Instrumentation & Measurement Magazine.

[100] B. Agathe, 2021, April. Novel anomaly detection and classification algorithms for IP and mobile networks, Research Thesis, https://hal.archives-ouvertes.fr/tel-03190474

[101] C.K. Reddy, B. Vinzamuri, A survey of partitional and hierarchical clustering algorithms, Data clustering: algorithms and applications: chapter 4, pp.88-107

[102] Retrieved from https://www.upgrad.com/blog/clustering-and-types-of-clustering-methods/

[103] S. Chitnis, P. Gokhale, August-2020, A Data Mining Framework using k-means to Analyze Highway Accident data in Maharashtra, International Journal of Management (IJM) © IAEME Publication Scopus Indexed, Volume 11, Issue 8, pp. 2058-2072, Scopus.com/sources.uri)

https://www.iaeme.com/MasterAdmin/uploadfolder/IJM_11_08_181/IJM_11_08_181.pdf

[104] k-means Clustering, from Wikipedia- The free encyclopedia article

[105] S. Haider, Q. Zhao, X. Li, 2020, Cluster-Based Prediction for Batteries in Data Centers, Center for Intelligent and Networked Systems (CFINS), Department of Automation and BNRist, Tsinghua University, Beijing 100084, China

[106] J. Hua, B. Lina, M. Wangb, J. Zhangb, W. Zhanga, Y. Lub, Health factor analysis and remaining useful life prediction for batteries based on a cross-cycle health factor clustering framework, Journal of Energy Storage, 50, 2022, pp. 1-10.

[107] N. Williard, W. He, C. Hendricks, M. Pecht, Lessons learned from the 787 dreamliner issue on lithium-ion battery reliability, Energies 6 (9) ,2013 4682–4695, https://doi.org/10.3390/en6094682.

[108] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, L. Zhang, Remaining useful life prediction for lithium-ion battery: a deep learning approach, IEEE Access 6 ,2018 50587–50598, https://doi.org/10.1109/ACCESS.2018.2858856.

[109] S. Tamilselvi, S. Gunasundari, N. Karuppiah, Abdul Razak RK, S. Madhusudan, V. M. Nagarajan, T. Sathish, M. Zubair M. Shamim, C. Ahamed Saleel, Asif Afzal, A Review on Battery Modelling Techniques, Sustainability MDPI, 2021, 13, 10042. https://doi.org/10.3390/su131810042.

[110] R. Borah, FR Hughson, J. Johnston, T. Nann. On battery materials and method, Material Today Advances, 6 ,2020, https://doi.org/10.1016/j.mtadv.2019.100046, pp. 1-13.

[111] J. Hong, D. Lee, E. Jeong, Y. Yi, Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning, Appl. Energy 278 ,2020, 115646, https://doi.org/10.1016/j.apenergy.2020.115646.

[112] L. LingLing, L. ZhiFeng, T. MingLang, A.S. Chiu, Enhancing the lithium-ion battery life predictability using a hybrid method - sciencedirect, Appl. Soft Comput. 74 ,2019, 110–121, https://doi.org/10.1016/j.asoc.2018.10.014.

[113] J. Li, M. Zhang, H. Zheng, J. Jie, Battery remaining useful life prediction using improved mutated particle filter, J. Energy Storage 3 (1) ,2021, e218, https://doi. org/10.1002/est2.218.

[114] S. Hong, T. Yue, H. Liu, Vehicle energy system active defense: a health assessment of lithium-ion batteries, Int. J. Intell. Syst. 2020 1–19, https://doi.org/10.1002/ int.22309.

[115] S. Hong, Y. Zeng, A health assessment framework of lithium-ion batteries for cyber defense, Appl. Soft Comput. 101 (49) 2021, 107067, https://doi.org/10.1016/j. asoc.2020.107067.

[116] R.R. Richardson, M.A. Osborne, D.A. Howey, Gaussian process regression for forecasting battery state of health, J. Power Sources 357 ,2017, 209–219, https:// doi.org/10.1016/j.jpowsour.2017.05.004.

[117] M.A. Patil, P. Tagade, K.S. Hariharan, S.M. Kolake, T. Song, T. Yeo, S. Doo, A novel multistage support vector machine based approach for Li ion battery remaining useful life estimation,

48

Appl. Energy 159 ,2015, 285–297, https://doi.org/10.1016/ j.apenergy.2015.08.119.

[118] R.R. Richardson, M.A. Osborne, D.A. Howey, Battery health prediction under generalized conditions using a Gaussian process transition model, J. Energy Storage 23 ,2019, 320–328, https://doi.org/10.1016/j.est.2019.03.022.

[119] S. Hong, J. Zhu, L.A. Braunstein, T. Zhao, Q. You, Cascading failure and recovery of spatially interdependent networks, J. Stat. Mech: Theory Exp. 2017 (10) ,2017, 103208, https://doi.org/10.1088/1742-5468/aa8c36.

[120] S. Hong, Z. Zhou, E. Zio, K. Hong, Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method, Digital Signal Process. 27 ,2014, 159–166, https://doi.org/10.1016/j.dsp.2013.12.010.

[121] S. Hong, Z. Zhou, E. Zio, W. Wang, An adaptive method for health trend prediction of rotating bearings, Digital Signal Process. 35 ,2014, 117–123, https://doi.org/ 10.1016/j.dsp.2014.08.006.

[122] M. Torabi, S. Hashemi, M.R. Saybani, S. Shamshirband, A. Mosavi, A hybrid clustering and classification technique for forecasting short-term energy consumption, Environ. Prog. Sustain. Energy 38 (1) ,2019, 66–76, https://doi.org/ 10.1002/ep.12934.

[123] I.P. Panapakidis, Clustering based day-ahead and hour-ahead bus load forecasting models, Int. J. Electr. Power Energy Syst. 80 ,2016, 171–178, https://doi.org/ 10.1016/j.ijepes.2016.01.035.

[124] C. XueJun, J. Zhao, J. XiaoZhong, L. ZhongLong, Multi-step wind speed forecast based on sample clustering and an optimized hybrid system, Renew. Energy 165 ,2021, 595–611, https://doi.org/10.1016/j.renene.2020.11.038.

[125] R. Schmuch, R. Wagner, G. Horpel, T. Placke, M. Winter, Performance and cost of materials for lithium-based rechargeable automotive batteries, Nat. Energy 3 (4) ,2018, 267–278, https://doi.org/10.1038/s41560-018-0107-2.

[126] L. Ungurean, G. Carstoiu, M.V. Micea, V. Groza, Battery state of health estimation: a structured review of models, methods and commercial devices, Int. J. Energy Res. 41 (2) ,2017, 151–181, https://doi.org/10.1002/er.3598.

49

Empirical Economics Letters, 22 (2): (August 2023) *ISSN 1681 8997*

An Approach to Enhance Sentiment Analysis on Social Media Data through Text Analytics and Predictive Modeling

Authors

Aditya Patil[a], Swati Patil[b], Prajakta Patil[b], Dr.Vaishali Patil[a], Sanket Lodha[c], Dr.Sudhir Chitnis[b], Dr. Arun Patil[b]

[a]Vishwakarma Institute of Information Technology, Kondhwa, Pune- 411048,India
[b]Vishwakarma College of Arts, Commerce and Science, Kondhwa, Pune-411048,India
[c]MGM University, Institute of Management & Research MGM Campus, N-6, Cidco, Aurangabad, Maharashtra 431003, India

Abstract

Understanding the feelings and thoughts of the general people has become increasingly important for businesses, government officials, and academics in this age of social media when millions of users are able to openly share their thoughts and experiences. Through the application of Natural Language Processing (NLP) and machine learning strategies, the present study intends to improve the sentiment analysis of data drawn from social media platforms. The text data will be pre-processed using NLP algorithms, which will allow for the extraction of essential features as well as information linked to sentiment.

Following this phase, models based on machine learning will be built to identify if the sentiment conveyed in the text is positive, negative, or neutral. The research will focus on adding sophisticated approaches for natural language processing (NLP), such as semantic analysis and sentiment lexicons, in order to improve the precision and specificity of sentiment categorization. In addition, the research will investigate the possibility of employing ensemble learning and deep learning algorithms in order to further improve the performance of sentiment analysis. Experimental assessments will be carried out using large-scale social media datasets in order to validate the effectiveness of the proposed strategy and compare it to other methods of sentiment analysis that are already in use.

The findings of this research will contribute to a more accurate sentiment analysis of the data collected from social media platforms. This will enable organizations to make decisions that are more informed and to have a deeper understanding of the sentiment held by the general population.

Keywords: Sentiment analysis, social media data, NLP, machine learning, data localization, feature extraction, sentiment lexicons

1. Introduction:

The continued use of digital social platforms has opened up several options for assessing interest in various aspects. These possibilities can be viewed not only from the point of view of individual users, but also from the point of view of groups of users. The rise in social media users has led to an explosion of content in many forms, including text, photos, audio, and video. Several hypotheses have been put out by researchers to account for the link between screen time and obesity [1][2]. An increase in people using social media is mostly responsible for this. This could lead to a lack of sleep, excessive eating throughout the scrolling, and extra pounds gained from commercial breaks. Inconsistent cross-sectional and prospective findings have been discovered in epidemiological studies of the relationship between screen time and less physical activity. This may be because of the difficulties in determining the extent to which people use screens and engage in physical activity.

However, individuals are able to openly discuss their thoughts, feelings, and experiences through the use of social media platforms, which have become a significant source of expressive content in the modern era. It is now absolutely necessary for a wide variety of stakeholders, including corporations, policymakers, and researchers, to get valuable insights from the vast amounts of data generated by social media.

Sentiment analysis, which comprises automatically identifying and classifying the sentiment that is expressed in text data, is one of the essential jobs in this sector. It is a method that employs machine learning, natural language processing, and computational linguistics to comprehend the emotions and viewpoints of social media users. Techniques based on natural language processing (NLP) and machine learning have proven to be quite helpful in improving the accuracy and effectiveness of sentiment analysis. Utilizing natural language processing (NLP) and machine learning techniques, the purpose of this study is to contribute to the advancement of the field of sentiment analysis on social media. Textual data will be handled by making strategic use of natural language processing (NLP) techniques, which will make it possible to derive important characteristics and information pertaining to sentiment from the data. After that, models of machine learning will be applied in order to categorize the sentiments as either positive, negative, or neutral.

51

The fundamental purpose of this research is to improve the precision and granularity of sentiment analysis by utilizing more complex natural language processing strategies such as semantic analysis and sentiment lexicons. In addition, cutting-edge methods such as ensemble learning and deep learning algorithms will be investigated to further improve sentiment analysis performance. The proposed research will be evaluated using vast, real-world social media datasets. This will make it possible for a full evaluation of the created methodology and its comparison against existing sentiment analysis techniques. The findings of this study will have far-reaching repercussions since they will enable organizations to obtain deeper insights into public mood and to make better-informed decisions based on data collected from social media.

2. Related work

The paper by P. Mehta et. al. [3] provides a comprehensive overview of the different modules of a sentiment analysis framework. The paper discusses the shortcomings associated with the existing methods or systems and proposes potential multidisciplinary application areas of sentiment analysis based on the contents of data. The strength of the paper is to provides a critical assessment of different modules of a sentiment analysis framework.The paper also proposes potential multidisciplinary application areas of sentiment analysis based on the contents of data. However, the paper does not provide a comprehensive picture of how to build a proper sentiment analysis model.The paper does not provide a detailed comparison of different sentiment analysis techniques and its limitations.

The paper contributes and provides a comprehensive overview of the different modules of a sentiment analysis framework and discusses the shortcomings associated with the existing methods or systems.

The foundations and scales of sentiment mining have received the bulk of academic attention. Authors depict many machine learning processes here. According to the study, Sentiment Analysis summarises the findings from many categorization techniques into positive, negative, and neutral ratings. The research demonstrates that machine learning techniques such as support vector machines (SVMs), naive bayes networks (NBNs), and neural networks (NNs), and in certain cases lexicon-based techniques, are highly accurate. Future work could involve figuring out how different sets of text data and other variables affect prediction precision. Therefore, it highlights the importance of continuing to work to enhance performance indicators in the future. Overall, the paper is a valuable resource for

52

researchers and practitioners interested in sentiment analysis and provides a foundation for future research in this area.

The research paper by J. Yadav [4] stated that sentiment analysis had three levels: sentence, document, and characteristics and categorised this information as positive, negative, or neutral. The study had used two primary approaches to sentiment analysis, which were machine learning and lexicon-based approaches. In this paper, social media mining and sentiment analysis were utilised to investigate social media information/data in consideration of a few leading companies across the world. Different companies from various parts of the world were considered for this paper and their data was analysed. Label details and storyboards were used to describe the information regarding sentiment analysis on various social media applications across the world.

Overall, according to this paper - sentiment analysis of tweets data could have helped to obtain valuable insights into the public's opinion on these topics, which would have allowed businesses, organisations, and decision-makers to make informed decisions based on the sentiment and the needs of their audience.

The paper entitled "Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques" by M. Kanakaraj and et. al. [7] is about the proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015) and focuses on the performance analysis of ensemble methods on Twitter sentiment analysis using NLP techniques. The paper explains the key idea of increasing the accuracy of sentiment classification by including natural language processing techniques, such as semantics and word sense disambiguation. It discusses the data gathering module, data processing module, training and classification module, and classification output of the proposed system. The document also compares the performance of ensemble methods against traditional ML algorithms and presents experimental results. The key points of the study include the use of ensemble classification to combine the effect of multiple machine learning algorithms, the analysis of sentiment from social network data, and the application of NLP techniques to mine sentiment from text data. The paper concludes that the ensemble method outperforms traditional classification methods, with extremely randomized trees classification performing the best. Overall, this research paper provides insights into the use of ensemble methods and NLP techniques for sentiment analysis on Twitter.

53

The book entitled, "Sentiment Analysis in Social Networks: A Machine Learning Perspective" by E. Fersini [8] provides a comprehensive overview of the state of the art in sentiment analysis, focusing on the challenges and opportunities presented by online social networks. The author discusses the nature of social networks, which are rich in informal languages and relationships among users. The book explores different models and approaches for sentiment analysis, including supervised, unsupervised, and semi-supervised models. It also highlights the importance of leveraging natural language and relationships in sentiment analysis and discusses the challenges posed by short and noisy content in social network messages. Additionally, it explores the future directions and potential applications of sentiment analysis in social networks. Overall, this book serves as a valuable resource for researchers, organizations, and institutions interested in sentiment analysis in the context of online social networks.

Pathak, R., & Waghmare, G. (2020) All marketing initiatives using the internet are included in digital marketing [9]. Through the use of search engines, social media, websites, and other platforms, digital marketing enables businesses to communicate with both present and potential customers. Knowing terms like retargeting and remarketing will help us better grasp digital marketing.

3. Proposed Method for analysis for on Social Media data

3.1 Sentiment Analysis

The methodology that will be proposed in this research will involve the use of a hybrid approach that will combine machine learning-based and lexicon-based methods. The machine learning-based method will make use of a labeled dataset to train the model, while the lexicon-based method is going to use a pre-defined set of sentiment lexicons to classify the sentiment of the text. The study will also suggest employing dimensionality reduction techniques to reduce the feature space and improve the performance of the model. The research will use a dataset of social media data available on Kaggle. Then it will evaluate the performance of the model using standard evaluation metrics such as precision, recall, and F1-score. The study is expected to find that the utilization of more advanced natural language processing strategies, such as semantic analysis and sentiment lexicons, will greatly enhance the accuracy and level of detail in sentiment analysis on social media data.

54

The lexicon-based strategy will count on a correlation between the presence of particular words or phrases in the text and an associated emotional tone. The studies will employ sentiment lexicons, which are collections of words that have been categorised according to their emotional connotations. The lexicons will serve as dictionaries, with numerous synonyms provided for each term. Manual methods, lexical methods, and corpus-based methods will all be used in the creation of the researchers' sentiment lexicon. Using the negative, neutral, and positive markers found in the lexicon, they will assign a polarity score to the provided text. Tagging words with a semantic orientation will be handled by the lexicon-based method, which can draw from dictionaries or corpora. Using a sentiment dictionary with opinion terms, we can simplify the former and easily determine the polarity score of words or phrases in the text. In order to boost the model's efficiency, the research article will employ a hybrid strategy that draws from both machine learning and lexicon-based techniques. Sentiment lexicons will greatly enhance the accuracy and granularity of social media sentiment analysis.

3.2 Machine Learning

Supervised, unsupervised, and reinforcement categories make up machine learning. Supervised Learning trains algorithms with labelled data. Regression and classification comprise Supervised Learning. Regression uses tagged data to produce continuous predictions. Regression issues have real-number output variables. This problem is usually linear. Data mining classification builds models using training data. This model classifies unknown class label records. Classification algorithms classify data. The AI must identify an object's categories and its confidence in its forecasts.

Second-type ML is unsupervised. Unlabeled data train the algorithm here. Clustering and dimensionality reduction comprise Unsupervised Learning. Clustering groups data sets in data mining. It involves dividing the population or data points into groups that are more similar than others. The goal is to cluster comparable groupings. K-means clustering algorithm is faster than others. It executes faster. Small and intermediate data work best. Our dataset is moderate. k-means is popular because it is 'easy' to implement algorithms. It can readily detect relevant feature relationships and extract multi-variate or composite variables. No clusters are needed. Elbow plot, implemented for accident data, can establish the optimal number of clusters. Dimensionality Reduction reduces data dimensions to remove unneeded data from input. It cleans data. ML has numerous dimensionality reduction algorithms, including PCA, LDA, and KPCA.

55

Machine Learning algorithms improve software prediction accuracy without programming [5][6]. Machine Learning can analyze such accidents. It is a fast-emerging trend that can help analyze patterns and uncover accident causes. Machine Learning can forecast accidents too. Learn from past data.

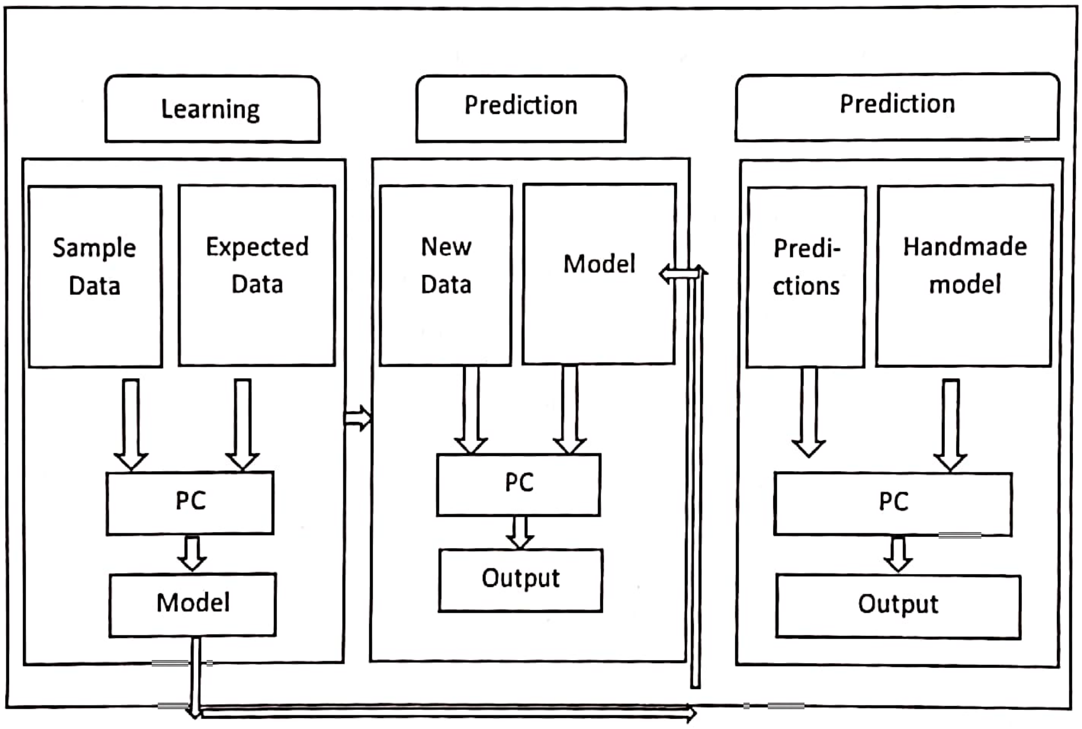

Figure 1: (a) Machine Learning Model Vs. (b) Traditional Model

Figure 1 (a) shows how a Machine Learning model differs from the Traditional model in Figure 1 (b). Unlike Traditional Modelling, Machine Learning Modeling's 'Learning' phase uses 'Model' and 'New Data' to forecast output. Learning and prediction are its two aspects. Pre-processing cleans data for learning.

The research will employ a hybrid strategy that combines machine learning and lexicon-based techniques. The former will utilise a labelled dataset to train the model, whereas the latter will use a predefined set of sentiment lexicons to classify the text's sentiment. The research will also employ dimensionality reduction techniques to reduce the feature space and enhance the model's performance. The approach based on machine learning will use supervised machine learning algorithms to classify the text's sentiment. In the study, the model's performance will be evaluated using standard evaluation metrics such as precision, recall, and F1-score. The results of the research will demonstrate that the proposed method will be more accurate and efficient than existing methods. This research will therefore employ predictive analytics and machine learning to enhance the precision and granularity of social media sentiment analysis.

Basically, understanding public perception of eco-friendly transformation and air quality, predicting box office success, and analyzing public sentiment during global events such as pandemics are all applications of sentiment analysis [4]. The purpose of sentiment analysis is to extract subjective information from text data. In order to identify and extract opinions, emotions, and attitudes conveyed in text data, natural language processing (NLP) techniques will be utilised. Analysis of social media, consumer feedback, and market research will increasingly incorporate sentiment analysis. The primary purpose of sentiment analysis is to categorise the tone of a given text as positive, negative, or neutral.

This hybrid approach strategy will combine methods based on machine learning and dictionaries to enhance the precision and granularity of sentiment analysis. The model will use a labelled dataset to train the machine learning-based method, while a predefined set of sentiment lexicons will be used to classify the text's sentiment using the lexicon-based method. Additionally, dimensionality reduction techniques will be utilised to reduce the feature space and enhance the efficacy of the model.

The machine learning-based technique will categorise the sentiment of the text using supervised machine learning algorithms including Support Vector Machines (SVM), Naive Bayes, and Random Forest. The lexicon-based approach will categorise the text's sentiment using a series of pre-defined sentiment lexicons. By compiling sentiment word lists utilising manual approaches, lexical approaches, and corpus-based approaches, the sentiment lexicons will be produced. In the study, the model's effectiveness will be assessed using common assessment measures like precision, recall, and F1-score. The study's findings will demonstrate that the suggested strategy will perform better than current approaches in terms of accuracy and efficiency. The following diagram will be used to demonstrate the hybrid technique utilised in the study so that readers will have a better understanding of the research process.

As a result, the study will use supervised machine learning algorithms, sentiment lexicons, and dimensionality reduction approaches to improve the model's performance. In terms of accuracy and efficiency, the proposed hybrid strategy may surpass existing methods. As a result, it could be a promising technique for sentiment analysis on social media data.

57

Sentiment Analysis Tool: This tool will perform sentiment analysis on user input and accurately determine the polarity of the expressed sentiments. The future proposed model will utilise data obtained from the localization database as its inputs. This tool will also utilise a range of machine learning techniques to generate highly accurate responses based on the given attitudes, using a suitable classifier. The proposed objective for this classification challenge is to develop a model that can automatically categorise attitudes into positive, negative, and neutral classes. This involves correctly assigning the suitable class label to a provided input. To utilise supervised classification, one must have a labelled text corpus for training, testing, and building the classifier.

Once we have acquired a labelled text corpus, the next important step is to develop a methodology for extracting features from the labeled corpus. This will be used to train the classifier. The effectiveness of the entire system relies on the quality of this specific feature extraction approach in the future proposed model. As a future proposed model, sentiment analysis will utilize various feature extraction techniques, such as unigrams, bigrams, and the incorporation of stop words.

Machine Learning: The utilization of the random forest algorithm in product analysing has been observed subsequent to a thorough comparison with the k-means clustering technique. The purpose of this comparison is to meticulously analyze prevailing trends and skilfully generate accurate forecasts. The random forest algorithm is exquisitely tailored to analyze seasonal data and generate exquisitely precise predictions for future values. The forthcoming illustrations elegantly showcase (Figure 3) the operational dynamics of this cutting-edge system, encompassing meticulous product analysis, trend analysis, and intelligent forecasting.

59

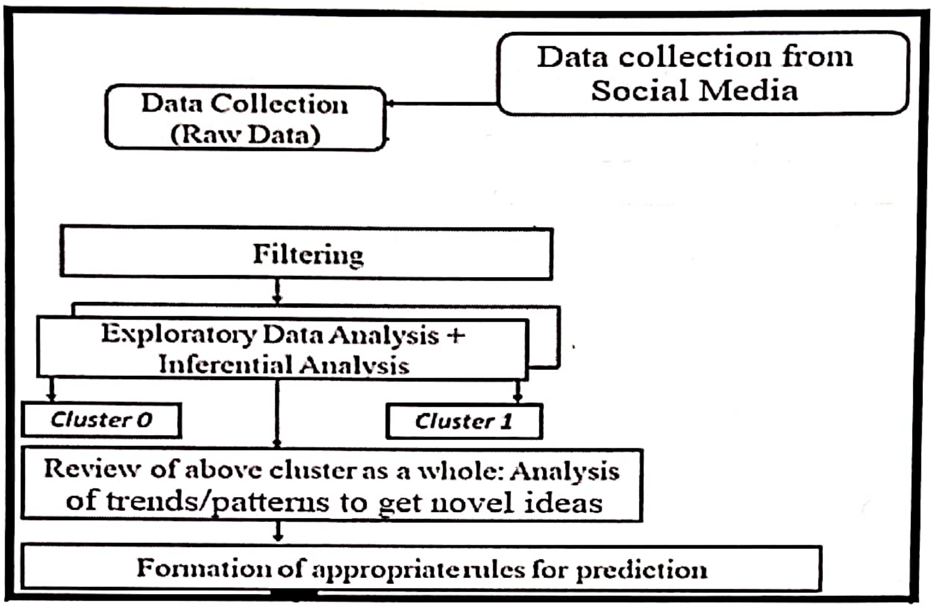

Figure 3:The operational dynamics of encompassing trend analysis and intelligent forecasting

Conclusion:

Analysing the cognitive processes of individuals within varying contexts and their perceptions towards diverse subjects will be a crucial undertaking. This will become increasingly significant in the context of the commercial domain, as businesses continue to depend heavily on their clients and consistently strive to develop products or services that will meet customer demands. Organizations will find more utility in understanding the preferences, opinions, and discussions surrounding current products, services, and brands when making decisions, such as competitor identification and trend analysis. The ability to communicate thoughts on social media platforms and access the resulting data will facilitate the aforementioned activities to some extent. The aforementioned initiative, titled "Sentiment Analysis for Social Media," will fulfil this purpose.

The proposed method for conducting sentiment analysis on the given dataset was dependent on the unique characteristics and attributes of the dataset used. The proposed method aims to investigate the scholarly aspects of the social media manifesto, focusing on its impact on the human dimension, knowledge dissemination, and circulation. This study seeks to explore the previously inconceivable or unattainable degree to which these factors are affected. Our proposed method for acquiring more insights involves focusing our research on the development of sentiment analysis models. The utilisation of these models played a crucial role in validating numerous hypotheses, ultimately aiding in the development of a comprehensive comprehension of the data and empowering society to take necessary actions.

The organised literature review will provide data on the analysis of sentiment studies that will be conducted on social media. Moreover, we will propose the method that will use to analyse the sentiments of people on social media. Many techniques will be given by the researchers, but one of the popular methods will be the Lexicon-based method with SentiWordnet and TF-IDF. In addition to that, machine learning techniques such as Naïve Bayes and SVM will be used.

Acknowledgement

We express our gratitude to the staff members and Dr. Arun Patil, the Principal-Director of VCACS-Pune, for their invaluable contributions, insightful comments, and assistance throughout our endeavour.

References:

[1] Strasburger VC; Council on Communications and Media . Children, adolescents, obesity, and the media. [published correction appears in Pediatrics. 2011;128(3):594]. Pediatrics. 2011;128(1):201–208 [PubMed] [Google Scholar]

[2] Robinson TN. Television viewing and childhood obesity. Pediatr Clin North Am. 2001;48(4):1017–1025 [PubMed] [Google Scholar]

[3] Mehta, P. Pandya,S.(2020) "A Review On Sentiment Analysis Methodologies, Practices And Applications", International Journal of Scientific and Technology Research, Volume 9, Issue 02, February 2020, ISSN 2277-8616.

[4] Yadav, J. (2023) 'Sentiment Analysis on Social Media', Qeios, CC-BY 4.0Article, January 9, 2023, Open Peer Review on Qeios, https://doi.org/10.32388/YF9X04.

[5] Sharma, D. Kumar, N.(2017, October). A Review on Machine Learning Algorithms, Tasks and Applications, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 6, Issue 10, pp. 1548-1552

[6] Park, C. Took, CC.Seong, J.(2018, February). Machine Learning in biomedical engineering, Biomedical Engineering Letters, Springer Link.

Empirical Economics Letters, 22 (2): (August 2023) *ISSN 1681 8997*

An investigation into the prognostic indicators for disease exacerbations with rheumatoid arthritis displaying subdued disease activity: an analytical approach

Authors
Prajkta Patil[1], Swati Patil[2], Sanket Lodha[3], Sudhir Chitnis[4]

[1]Prajakta Patil VCACS
[2]Swati Patil VCACS
[3]Sanket Lodha MGM University, IOMR
[4]Dr.Sudhir Chitnis VCACS

Abstract

Rheumatoid arthritis (RA) is a prevalent and enduring inflammatory disorder that exerts its impact on a substantial global population. It is a persistent autoimmune condition characterised by the presence of joint inflammation and pain. Patients diagnosed with RA frequently encounter episodes characterised by diminished disease activity, wherein their symptoms manifest in a mild or even non-existent manner. Nevertheless, it is imperative to acknowledge that these periods may be subject to interruption by flares, wherein individuals endure an abrupt escalation in disease activity and corresponding symptoms. Flares, in their devastating nature, possess the capacity to exert a substantial influence on the overall quality of life experienced by patients.

The primary objective of this study is to distinguish the prognostic indicators that may be associated with disease exacerbations in patients diagnosed with RA who demonstrate low disease activity. The present investigation aims to conduct a comprehensive analysis of data obtained from a cohort of RA patients exhibiting low disease activity. Proposingadvanced machine learning methodologies, our objective is to discern the key factors that serve as predictors for the occurrence of flares in this particular population. The study will additionally explore the intricate interplay between these factors and various clinical outcomes, including but not limited to quality of life and radiographic progression.

The findings from proposed methodology derived from this study will yield invaluable insights pertaining to the effective management of patients diagnosed with RA exhibiting low disease activity. Moreover, these findings will serve as a valuable tool for clinicians, enabling them to accurately identify patients who are susceptible to experiencing disease flares. By virtue of promptly identifying these patients and administering suitable

therapeutic interventions, clinicians have the capacity to avert exacerbations and enhance the overall quality of life for patients.

Keywords: RA, flares, exacerbations, machine learning, autoimmune, joint inflammation

Introduction:

RA is a chronic autoimmune condition characterized by the presence of joint inflammation and the accompanying sensation of pain. Inflammatory arthritis (IA) refers to the joint inflammation that arises due to an immune system that is excessively active. Typically, this condition exhibits a simultaneous impact on numerous joints inside the body, although it may also manifest in a singular joint. Inflammatory forms of arthritis exhibit a lower prevalence compared to osteoarthritis (OA), which stands as the most prevalent form of arthritis. However, RA is commonly considered a condition that encompasses multiple diseases [7], each of which is attributed to distinct pathogenic pathways. Hence, it is logical to anticipate that no singular rheumatoid arthritis (RA) model would fully replicate all facets of the human disease. Nonetheless, the diverse animal models available can be employed to elucidate distinct pathways that contribute to the development of arthritis. Individuals diagnosed with RA commonly encounter fluctuations in their symptomatology, characterized by intermittent exacerbations known as flares. Flares, in their unpredictable and debilitating nature, pose a significant challenge for medical professionals in establishing a universally accepted standard definition to effectively guide their treatment approaches [2][6]. Prognostic factors play a crucial role in guiding treatment decisions for rheumatoid arthritis (RA). Among the various factors considered, those that indicate high disease activity, early presence of erosions, and autoantibody positivity are commonly recognised as poor prognostic indicators [1][5]. In addition to the aforementioned factors, it is imperative to evaluate other aspects such as functional disability, extraarticular disease, and multi-biomarkers [1]. The primary aim of this research study is to delve into the prognostic indicators associated with disease exacerbations among patients diagnosed with rheumatoid arthritis (RA) who exhibit subdued disease activity [1]. Rheumatoid factor (RF) and total immunoglobulin isotypes as well as serum levels of soluble interleukin-2 receptor (sIL2-2R) were determined by the Margreet Kloppenburg and et. al., [3] in order to study immunological parameters of the disease, whereas patients with rheumatoid arthritis-associated interstitial lung disease (RA-ILD) or idiopathic pulmonary fibrosis (IPF) could have an unexpected acute exacerbation

63

ANVESAK
ISSN: 0378-4568
UGC Care Group-I Journals

RURAL NON-AGRICULTURAL ENTREPRENEURSHIP: CONSTRAINTS AND OPPORTUNITIES

Dr. Shital Mantri

MBA-HRM & Marketing, PhD, HOD, Vishwakarma College of Arts, Commerce & Science, Kondhwa, Pune-411048, INDIA

Dr. Sheetal Waghmare

MBA-HRM, PhD, Vishwakarma College of Arts, Commerce & Science, Kondhwa, Pune-411048, INDIA

Vaishali Kale

MCA, Set, Coordinator BBA (CA), Vishwakarma College of Arts, Commerce & Science, Kondhwa, Pune-411048, INDIA

Abstract

The rural sector has long been thought of as one that produces low-quality, low-productivity commodities. It is frequently predicted that it will disappear as a nation advances. The acknowledgment that the rural non-farm sector may and frequently does contribute to economic growth, rural employment, poverty reduction, and a more spatially balanced population distribution has emerged in recent years as a departure from this position. A rural entrepreneur is a person who frequently has a wide range of possibilities when deciding which areas of operations to pursue. In general, a business can be established in any one of the primary (agricultural), secondary (industrial), or tertiary (service) sectors of the economy. One of the most crucial factors in the economic development of a nation and its regions is the rural entrepreneur. India has historically been an agricultural nation. Alongside its traditional sources of income, such as agriculture, traditional handlooms, and handicrafts like weaving and pottery, rural India's non-agricultural sectors are subtly growing. The economic contribution of agriculture to India's GDP has been declining over time. Historically, the non-agricultural rural sector has been viewed as a low-productivity, low-quality sector. The acknowledgment that the rural non-agricultural sector may, and frequently does, contribute to economic growth, rural employment, poverty reduction, and a more geographically balanced population distribution has emerged as a departure from this position. Therefore, lowering the rural populations' reliance on agriculture as a source of income will contribute to raising their overall standard of living. The identification of numerous issues relating to rural non-agricultural entrepreneurship is the main topic of this research. The best ways to solve these issues have also been considered.

Keywords: Economic development, Non-agricultural rural entrepreneurship, problems of Non-agricultural.

Introduction

Despite the continuous development of cities all over the world, in 1950, more than 79% of the population was still living in villages. In India in 1961, 82% of people lived in rural areas, whereas in 1971 the proportion was 80%. Obviously, rural development is very necessary to achieve national development through economic planning and therefore efforts should be made to understand the nature of rural problems and solve them. In any nation, the rural society plays a remarkable role. Food grains and other raw materials are produced in the rural areas and the needs of the cities are met only through the rural production. Moreover, the rural areas are also responsible for providing labor to the industrial businesses in the cities. Most of the nation's natural resources are available in rural areas and most of the population resides there. We know relatively little about the function of the rural non-agricultural sector in the broader development process and it is a poorly understood part of the rural

64

economy of emerging nations. Due to the industry's extreme variability and insufficient attention at the theoretical and empirical levels, there is a knowledge gap in this area. Many people hold the belief that rural off-farm employment is a low productivity industry that produces low quality goods and would eventually disappear as a nation improves and incomes rise. The main justifications for paying attention to the non-farm sector are its alleged capacity to absorb a growing rural labour force, to limit rural-urban migration, to contribute to the rise of the national economy, and to support a more fair distribution of income. Whether non-agricultural activity is more efficient than its urban or agricultural counterparts at converting resources into output is crucial when evaluating its capacity to contribute to development. Cities have fueled the growth of industries in India over the last five decades. As a result, people from rural areas flocked to cities in large numbers in search of work. Urban slums expanded, and poverty persisted, as did the issue of rural development. Even now, 70% of the population of the nation still relies on agriculture as their primary source of income. However, the climate fluctuated and agriculture was dispersed. A criticism is made that agriculture is insufficient. Therefore, rural entrepreneurship is required in addition to agriculture in order to deliver financial benefits from agricultural products to the rural areas.

Objective of Study

1. To investigate the economic development functions of rural non-agricultural entrepreneurs.
2. To understand the importance of rural non-agricultural entrepreneurship.
3. To investigate the primary problems confronting rural non-agricultural entrepreneurship in the Pune District.
4. To propose some of the primary solutions to difficulties in rural entrepreneurship.

Limitations

The study is limited to Ambegaon Taluka, Pune District in Maharashtra only.

The researcher has selected only few non-agricultural Enterprises.

Research Methodology

The research paper is based on both Primary and Secondary Data. Primary data is collected after visiting the markets of Ambegaon Taluka of Pune district in Maharashtra, India. Interview Schedule method is used to elicit information from the Entrepreneur who were involved in various small businesses like oil mills, seeds and fertilizers shop, dairy farm, cement and brick business, wholesale vegetable and fruits vendors etc. In order to compact the study 50 Entrepreneurs were selected on the basis of Random Sampling Method. Secondary data collected from the Commerce and Management Books, Journals, Magazines published.

Literature Review

Today's young generation is more entrepreneurial than the previous one. People are going out on their own in increasing numbers. Rural entrepreneurship created significant amount of potential in rural youth. Entrepreneurship is not limited to urban areas only, it has created pavement for the rural sectors as well. Studies around the world have proposed the opinion that small-scale rural non-agricultural entrepreneurs can establish their business in better way. It has emerged that small-scale entrepreneurs have long back given up the idea of following traditional methods of doing business. Better alternative technologies aimed at increasing productivity, diversifying production, reducing risks, and ultimately increasing profits. Today's young generation is more entrepreneurial then the previous one. People are going out on their own in increasing numbers. Rural entrepreneurship created significant amount of potential in rural youth. Entrepreneurship is not limited to urban areas only, it has created pavement for the rural sectors as well. Studies around the world have proposed the opinion that small-scale rural non-agricultural entrepreneurs can establish their business in better way. In essence, small-scale non-agricultural rural entrepreneurs are learning to take calculated risks to develop or create new markets for their products as they become more market-oriented. It has been proven thus far that innovation is

65

key to entrepreneurship and that many other crucial resources are required in addition to invention. Rural businesses have become more imperative to the economy. (Banarjee G.D.,2011). Banarjee claims that "high degrees of integration lead to greater economic activity, which not only creates jobs for locals but also helps society as a whole." Small-scale businesses with high growth potential contribute to rural development. It is also stated that despite exceedingly harsh and costly market circumstances, small-scale firms in India have flourished. It is also argued that small-scale businesses in India have survived in extremely unfavorable, expensive market conditions. Yet, small enterprises in rural regions confront several problems, including a lack of qualified personnel, restricted space availability, a lack of technology and managerial innovation, as well as, to some extent, remoteness from urban areas (Keeble, 1993). On the other hand, other authors argue that the issues of operating a small company are the same regardless of location. The promotion of entrepreneurship in rural areas is intended to pave the way for the abolition of poverty. Rural business owners, however, are under informed and have limited access to resources and infrastructure. Rural entrepreneurs have very limited access to financial, educational, and training resources, which limits their ability to launch new businesses and achieve commercial success. Entrepreneurs are constrained regarding access to markets, and they are significantly dependent on middlemen who exploit rural entrepreneurs by making a high profit. Their lower educational and training background has also affected their entry to business, and it has also affected the acquisition and procurement of raw materials as it pushes them to procure with a poor quality of raw material. Non-agriculture enterprises require entrepreneurial abilities to start, expand, and sustain the business. Though non-agricultural enterprises are mostly microenterprises and informal businesses that require little capital, the skills required maintaining and managing, and build the business are significant.

Importance of Rural Non-Agricultural Sector

In recent years, the non-agricultural sector, particularly in rural areas, has received widespread recognition for the following reasons:

Check for Unrestricted Migration: A planned strategy of non- agricultural development in rural areas could stop many rural residents from moving to metropolitan commercial and industrial hubs.

Overcoming the urban-rural divide Rural-urban economic divides: This will inevitably close as the rural economy diversifies its economic foundation beyond agriculture, which will also have positive implications on many other facets of human life and ambitions.

Employment Creation: Non- agricultural rural sectors tend to be more labor-absorbing and less capital-intensive. They may therefore be extremely important to the expansion of employment in rural areas.

Reducing Inequalities: Rural communities with a large network of non-farm employment opportunities have significantly less unequal income distribution. Social mobility is aided by the lower socioeconomic strata of rural communities participating much more actively in non-agricultural activities.

Role of Rural Entrepreneurs in Economic Development on Economic Development:

Entrepreneurs have a talent for analysing, identifying, and scanning their environment for possibilities, which they then convert into workable business concepts by establishing legitimate firms. By shifting resources from less productive to more productive uses, they generate wealth. By utilising the nation's resources effectively and efficiently, they act as catalysts for both economic growth and as forces for social revolution.

An entrepreneur creates an industry and hires lots of people. In addition to giving the government money in the form of taxes to operate the entire social system, the society also supports many people's ability to make a livelihood through economic transactions, either directly or indirectly.

1. Entrepreneurship Speeds Up Economic Growth: By developing new goods and services, entrepreneurs drive the creation of new jobs, which in turn speeds up economic growth. Therefore, it

66

makes sense to view public policy that encourages entrepreneurship as essential to the advancement of the economy. Numerous new opportunities and occupations are created by entrepreneurship. Entrepreneurship generates a sizable number of entry-level jobs, which are essential for converting unskilled individuals into competent ones. Additionally, it creates trained personnel to perform jobs in large companies.
The rise in a country's overall employment rate is mostly attributable to the expansion of entrepreneurship. Thus, entrepreneurship plays a significant role in the creation of new job opportunities.

2. The creation of jobs: Rural industries give rural residents several job options. Through rural industrialization, the fundamental issues of widespread underemployment and unemployment in Indian Population can be effectively addressed.

3. Promotes economic growth: Rural industrialization encourages the economic growth of rural communities. On the one hand, this slows rural-to-urban migration, and on the other, it lessens the uneven rise of cities and towns the expansion of slums, social unrest, ecological degradation, etc.

4. Foreign exchange earnings: Through the export of their products, rural industries contribute significantly to the country's increased foreign exchange earnings.

5. Produces goods of the consumer's preference: Rural industries, especially village and cottage industries, generate goods of the preference and taste of the consumer. Jewelry, saris, and other artistic objects are made to meet the needs of various consumers based on their preferences in taste, style, and design.

6. Development of entrepreneurship: Entrepreneurship is encouraged in the agricultural area through rural industries. It promotes the creation and execution of entrepreneurship development in the rural sector, thereby facilitating the development of rural areas.

Problems faced by Rural Non-agricultural entrepreneurs

The advancement of the economy and subsequent increase in the country's Gross Domestic Product are both undeniably greatly influenced by entrepreneurs. In their everyday lives and at work, they deal with a wide range of difficulties. Everyone is aware that owning a business has its drawbacks and that even successful companies can encounter difficulties. Among the most pressing problems faced by rural businesses are the following:

1. Financial Problems:
Due to a glaring lack of security and credit in the market, the majority of rural businesses are unable to secure outside funding. 46 out of the 50 business owners responded, accounting for 92% of the respondents. They think that bank loan applications take too long and frequently irritate rural company owners. One of the main issues rural businesses currently face, especially in light of the global recession, is the lack of available financing. Because it is expensive to go to banks and there are other obstacles in the way, some nationalized banks have not attempted to reach out to rural consumers. Since they frequently lack adequate established forms of collateral or tangible assets, poor individuals are frequently excluded from the regular financial market.

2. Marketing Problems:

a. Competition from big businesses and metropolitan business people is fierce for rural entrepreneurs. Due of the high input prices, they have expensive manufacturing expenses. Regulations and the pressure from large scale units on competition, according to nearly 76% of respondents, or 38 persons, are major problems for marketers. Entrepreneurs lament how hard it is to establish standards and then uphold them.
b. Middlemen: Middlemen take advantage of rural business owners. The vast majority of the study's entrepreneur respondents note that they rely extensively on intermediaries to sell their goods, which cause them unnecessary hassle and take a sizable portion of the profit as commission. According to 34

out of 50 respondents (68%), middlemen cause problems for their firm. Other marketing issues in remote locations include inadequate storage and transportation options. In communities, farmers often store their food in open spaces, clay buckets, sacks, etc. Hence, these native and conventional techniques of storage are unable to shield the crop from moisture and other threats. Unfortunately, agricultural products are neither standardized nor rated.

3. Lack of familiarity with information technology: In general, information technology is not particularly prevalent in rural regions. Entrepreneurs depend on IT facilities because of the lockout and other unfriendly circumstances. The enterprises of entrepreneurs are being hampered by inadequate internet capabilities and an absence of technological understanding. 76% of the respondents, or 38 people, stated they lacked the technical and IT skills necessary to execute their jobs.

4. Dearth of technical knowledge: Technical ignorance is a significant barrier for rural businesses. A barrier to the growth of rural entrepreneurship is the absence of the necessary training facilities for managing a small firm and other extended services. of 50 people surveyed, 25%, or 50%, indicated they had no formal education.

5. Workers with low skill levels: Most rural businesses struggle to locate staff with high skill levels. The workers' dropout rate is also significant in this situation. Workers need to receive on-the-job training, and it has to be delivered in a language they can readily grasp.

6. Procurement of Raw Materials: One of the most challenging jobs for a rural entrepreneur is sourcing raw materials. A total of 36 respondents, or 75% of the total, expressed concern that they would get subpar raw materials and run into storage and warehousing issues.

7. Traditional Families are a Problem: Traditional and strict households forbid the younger generation from engaging in novel businesses or endeavors. Of 50 respondents, 45 (or 90%) said that traditional families are against women entrepreneurs starting businesses because it goes against their values and traditions. 8. Policy Challenges: 80% of the respondents, who make up the majority, assert that the atmosphere is not favored by government policies for new entrepreneurs. Rural business owners do not receive a sizable quantity of government subsidies or freebies from the relevant agencies. The working young entrepreneurs find this annoying.

a). Political and structural problems: Before establishing the business, entrepreneurs clear the government complicated like business license, pollution and clearance etc. Due to low level of education rural entrepreneurs may not complete this process fastly.

Suggestions

There various governmental and non-governmental institutions are taking efforts to sole the difficulties faced by the rural non-agricultural entrepreneurs. Solutions relating to pricing of the products, distribution channels to the rural entrepreneurs, packaging and labeling, advertising are all some issues relating to marketing issues. To promote rural entrepreneurs to create firms, the following policies could be implemented:

1. Making availability of loans at low interest rates: Rural businesses should have access to financing with flexible repayment periods and low interest rates. Moreover, lengthy procedures should be avoided when financing rural business owners.

2. Facilitating Finance Cells: Banks and financial institutions must establish special finance cells to provide simple financing to rural entrepreneurs.

3. Creation of marketing cooperatives: It is important to support and motivate rural entrepreneurs who want to create marketing cooperatives. Cooperatives may undoubtedly help in both procuring inputs at cheap rates and helping to sell their goods at acceptable costs. Rural Business owners may benefit from business by cutting out intermediaries. Joint production and marketing hubs with cutting-edge modern infrastructure should also be built to support rural enterprises. As a result, robust training infrastructure, high-quality instruction, the expansion of marketing cooperatives, and the creation of unique financial institutions may all help rural companies in India thrive. Governmental and non-governmental organizations should both play a significant role. Ensuring an adequate supply of rare raw materials for rural business owners should be a top priority. In addition to subsidies, it must be possible to give prices that are both reasonable and competitive for the goods produced by rural businesses.

4. Supplying top-notch facilities: Training is essential for the development of enterprise. Rural entrepreneurs can effectively establish their business because proper training equips them with the required skills. The Prime Ministers rozgar yojna (PMRY), Rotary Clubs, Lion Clubs, and other nonprofit organisations that can also organised these development programs for rural entrepreneurs to give them proper assistance and motivational counseling are currently providing such training facilities to economically weaker entrepreneurs. Because they combine motivational and familiarization processes with the potential of different credit facilities and support through escort services, individual-based EDI approaches are highly relevant for the development of entrepreneurship. They may be used to persuade rural youth with the necessary fundamental skills and hands-on future technologies to launch small businesses in their area.

Conclusion

Increasing accessibility to the community and closing gaps in specific areas, such as providing industry experts as mentors, setting up incubation centers, and holding ideation workshops, hackathons, and other events in non-metro cities to foster innovation and entrepreneurial mindsets, are all ways to fundamentally promote rural entrepreneurship. While the Indian government is very important, NGOs and other civil society organisations play a significant role in institutionalizing support systems. Rural businesses are essential to the development of our nation, but they face several obstacles, including limited access to financing and inadequate infrastructure. Even though the government has taken action to address the problems, more concerted efforts are required. With the appropriate assistance, access to training, and funding,

References

1. Chaudhary Garima, : A Study on Marketing Aspect of Medium and Small Enterprises, International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-3, Issue-10)
2. Economist. (2009). Reforming through the Tough Times, September 12, pp. 71.
3. Goel, A., Vohra, N., Zhang, L., & Arora, B. (2007). Attitudes of the Youth Towards Entrepreneurs and Entrepreneurship: A Cross-Cultural Comparison of India and China. Journal of Asia Entrepreneurship and Sustainability, 3(1), pp. 1-35
4. Gupta, V. (2009). Cross-Cultural Similarities and Differences in Characteristics Attributed to Entrepreneurs: A Three-Nation Study. Journal of Leadership & Organizational Studies, 15(3), pp. 304-318.
5. Kotler, P., Kevin, K., Koshy, A., & Jha, M. (2011). Marketing Management (13th ed.). New Delhi: Pearson Publications.
6. Kshetri, N. (2011). The Indian Environment for Entrepreneurship and Small Business Development. Studia Negotia, 56(4), pp. 35-52. 7. R.S.Kanchana*, J.V.Divya and A.Ansalna Beegom, Challenges faced by new entrepreneurs, ISSN: 2347-3215 Volume 1 Number 3 (2013) pp.
7. 71-78 8. Entrepreneurship: Problems and Challenges Faced Ajay Nayar, Vasanth Kiran.

SARDAR PATEL INSTITUTE OF ECONOMIC AND SOCIAL RESEARCH

anveṣak

A bi-annual journal

CERTIFICATE OF PUBLICATION

This is to certify that the paper entitled

RURAL NON-AGRICULTURAL ENTREPRENEURSHIP: CONSTRAINTS AND OPPORTUNITIES

Authored by

Dr. Sheetal Waghmare

UGC

University Grants Commission

Approved Journal

vol. 53 No. 01

in

Anvesak A bi-annual Journal

UGC Care Group - 1

ISSN: 0378-4568

January-June 2023

Impact Factor: 6.20

Anvesak A bi-annual Journal

Editor

Anvesak

A bi-lingual Journal

Vishwakarma College of Arts Commerce & Science

ID No PU/PN/ACS/ 275/2007

College Code No. 824

Pune-48

SARDAR PATEL INSTITUTE OF
ECONOMIC AND SOCIAL RESEARCH

anveṣak

A bi-annual journal

CERTIFICATE OF PUBLICATION

This is to certify that the paper entitled

RURAL NON-AGRICULTURAL ENTREPRENEURSHIP: CONSTRAINTS AND OPPORTUNITIES

Authored by

Dr. Shital Mantri

UGC

University Grants Commission

Approved Journal

vol. 53 No. 01

in

Anvesak A bi-annual Journal

UGC Care Group - 1

ISSN: 0378-4568

January-June 2023

Impact Factor: 6.20

Anvesak A bi-annual Journal

Editor

Anvesak

A bi-lingual Journal

Vishwakarma College of Arts Commerce & Science * Pune-43 *
ID No. PU/PN/ACS/ 275/2007
College Code No. 324

SARDAR PATEL INSTITUTE OF ECONOMIC AND SOCIAL RESEARCH

anveṣak

A bi-annual journal

CERTIFICATE OF PUBLICATION

This is to certify that the paper entitled

RURAL NON-AGRICULTURAL ENTREPRENEURSHIP: CONSTRAINTS AND OPPORTUNITIES

Authored by

Ms.Vaishali Kale

vol. 53 No. 01

in

Anvesak A bi-annual Journal

UGC Care Group - 1

ISSN: 0378-4568

January-June 2023

Impact Factor: 6.20

Editor

Anvesak

A bi-lingual Journal

Vishwakarma College of Arts Commerce & Sci. ID No. PU/PN/ACS/ 275/2007 College Code No 324 * Pune-48 *

ANVESAK
ISSN: 0378-4568

UGC Care Group-I Journals

ISSUES AND POSSIBILITIES IN MARKETING OF AGRICULTURAL COMMODITIES

Poonam Jadhav

Assistant Professor, Vishwakarma College of Arts, Commerce & Science, Kondhwa, Pune

Dr. Shubhangi Auti

Professor, Annasaheb Magar Mahavidyalaya, Hadapsar, Pune

Abstract

The study discusses the challenges and opportunities of rural marketing in India. The rural market in the Indian economy is divided into two basic divisions. Rural markets have grown in importance in recent years, as economic expansion has resulted in a significant increase in rural people's purchasing power, and rural people's preferences are changing. As a result, every marketing player wants to invest in rural areas. Though there are enormous potential and large growths opportunities in rural markets, there are several issues that have created barriers to accessing rural markets. This study advances the exploration of numerous rural marketing strategies as well as the existing rural marketing ecosystem, highlighting major difficulties and suggestions linked to rural marketing.

Keyword: Rural marketing, marketing issues in rural areas, rural marketing potentials

Introduction:

Agricultural marketing is broadly the exchange or bartering of agricultural produce. To form such exchanges possible various processes viz., processing, storage, transportation, grading, inspection, pricing, advertising, wholesale and retail sale etc are included. From the purpose of view of the govt. the function of Agriculture Marketing is to link the assembly of agricultural commodities with sustainable supply and trade that's economically beneficial. At the government level, the government can provide services to the farmers by fixing marketing intelligence, i.e. the market value of agricultural products and a system that builds barriers in terms of cultivation. Government can encourage farmers by creating policies to make different options regarding supply chain. Farmers employed farm sector inputs like local seeds and farmyard manure within the past. These inputs were easily accessible to them; farmers' market purchases of crop inputs were low. Farm inputs, like improved seeds, fertilizers, insecticides and pesticides, farm machinery, implements, and financing are increasingly important within the production of farm products in recent years. The new farming technology responds to input. Product and input marketing must be included within the scope of agricultural marketing.

Objective of Study

Objective of Study Rural markets offer unrealized potential as a component of any economy. Several challenges confront the search to thoroughly explore rural markets. The concept of agricultural markets in India, also as in several other countries like China, remains growing, and therefore the sector presents variety of issues, including understanding the dynamics of rural markets and developing methods to provide and satisfy rural consumers.

Research methodology

This research article attempts to provide a deeper understanding of the function of rural marketing in economic development. The study also intends to investigate the issues confronting rural marketing in the current context. The descriptive research approach is utilised in this study for this goal, which is based on the utilisation of secondary sources of data acquired from books, Journals, periodical publications, government publications, articles, newspapers, and websites, among other things.

Rural marketing concept

Rural marketing in the Indian economy can be broadly divided into two categories.

The marketplaces for consumer durables include both durable and non durable items. The agricultural product markets, which include those for seeds, insecticides, fertilizers, and other items.

People who believe rural marketing in India to be exclusively about agriculture marketing occasionally make this error. Rural marketing influences how economic operations are transported from metropolitan areas to rural areas as well as how different products made by non-agricultural employees are marketed from rural to urban areas.

Definitions of Rural Marketing:

Identifying the needs of customers and potential customers, providing products/services that satisfy their needs, and developing efficient processes or systems to deliver your product/service to the market when, where, and how consumers want it.

The following are the characteristics of rural markets –

Here, agriculture is both the first and primary source of income, and this revenue is cyclical and variable because it is based on crop yields.

- The rural market, despite its size, is dispersed geographically.
- Disparities in religion, culture, and economic status are evident.
- The market isn't too developed because the locals have enough money to buy things.
- It exhibits sharper and diverse regional preferences with distinct forecasts, habit patterns, and behavioral features.
- These markets have their orientation on agriculture, with a poor standard of living, low per capita income, and backwardness.
- The overall rural development process leads to the rural marketing process. The core of the rural marketing process is the inception and administration of social and economic change in the rural sector.

Issues faced by Rural Market

Various obstacles prevent the rural market from developing. When marketing to rural areas, marketers confront a variety of issues including physical distribution, logistics, a lack of a proper and effective sales staff, and a lack of an effective marketing communication system.

The following are the main issues that rural markets face:

- **Numerous Middlemen**

Agricultural marketing is considered as a complex process that involves a large number of intermediaries managing a wide range of agricultural goods that are distinguished by seasonality, bulkiness, highly perishable, and so on. The predominance of these middlemen varies with commodities and product marketing channels. The producer's portion of the consumer's burr (pocket) is lowered as a consequence of the involvement of several middlemen.

- **Small and Scattered Holding**

Due to the small and dispersed size of the agricultural holdings, there is a small amount of marketable surplus produced. It is not an easy task organizing how the products can be assembled for efficient marketing. Moreover there are many varieties of particular crops such as millets and this poses problems in pricing.

However, there would be minimally bureaucratic storage procedures. The depositor who wishes to store the produce in the warehouse must submit a written request in the format required by the warehouse. The item intended for storage will be dutifully delivered to the warehouse in a properly packaged state. In the application form, the depositor must provide all information about the commodity, including its market worth. Before being stored, the commodity brought in for storage will be graded and weighed by trained technical experts. For other commodities, separate storage fees would also be applicable, and the stocks available for storage would be covered by insurance against potential dangers of fire, theft, and floods, strikes, and civil unrest.

Administration of the Sales Force

In order to address the problems with sales force management, the company pays special attention when employing and selecting salespeople since the qualities they need differ from those required by salespeople in cities. When working with rural customers, these salespeople need to be patient and fluent in the regional or local tongue.

Managing such a large and distributed sales staff, overseeing their sales calls, helping them with their personal and professional concerns, and motivating them to achieve better results should be both exciting and hard for the sales manager. Therefore, persons who work in rural areas should have a natural desire to assist the community and build ties with its residents.

Information on the Market

As a result, we have publications like newspapers, price bulletins, government agency reports, etc. that offer market information. If a programme for analysis and interpretation of market data was made available, it would make this information considerably more helpful. The farmers can make use of the raw data thanks to a professional interpretation, which undoubtedly yields significant information.

Suggestions for Effective Rural Market

The study of consumer demand by a business in order to increase output and promote a product is known as market research. It focuses on consumer demands, preferences, product impressions, market accessibility, marketing effectiveness, etc. The marketing system needs to be improved, and market research must be given top importance.

Suggestion

- There should be better marketing information system development in India
- The Retail sector should grow as per the requirements of Agricultural Marketing
- Government should develop unique marketing strategies for agricultural sector in India

India has to enhance its marketing information systems more effectively. The retail industry should expand in accordance with the needs of agricultural marketing. The government should create distinctive marketing plans for India's agricultural industry.

Conclusion

In India, agricultural marketing is expanding quickly. This may represent a special area of economic advantage for India. The nation's economy would steadily grow if agricultural marketing were to improve. In India, agricultural marketing is becoming more prevalent throughout all states. Additionally, agricultural marketing places a great value on this.

References

1. Dogra B, Ghuman K, Rural Marketing: Dogra B, Ghuman K, Rural Marketing Concepts & Practices, Tata McGraw-Hill, 2012.
2. Mr. K. Phanindra Kumar, Mr. S. Swamy," Indian Rural Market – Opportunities and Challenges".

3. Shabana Siddique, R. A. Siddique, "Rural Marketing in India: Strategies".
4. Mr. K. Phanindra Kumar, Mr. S. Swamy, "Indian Rural Market — Opportunities and Challenges"
5. Shabana Siddique, R. A. Siddique, "Rural Marketing in India: Opportunities, Challenges And Strategies"
6. Kohls, R.L. and Uhl, J.N. (1990). Marketing Of Agricultural Products, 6th edition, Macmillan Publishing Company, New York, p.355
7. Kohls, R. L. and Uhl, J. N. 2002. Marketing of Agricultural Products. Ninth Edition. Prentice Hall, chapters 23-29.
8. Kotler, P. and Keller, K. L. 2006. Marketing Management. Twelfth edition, Prentice Hall, Englewood Cliffs, N.J., chapters 1-2.
9. Retrieved from http://en.wikipedia.org

77

M.C.E. Society's

ABEDA INAMDAR SENIOR COLLEGE

OF ARTS, SCIENCE & COMMERCE (AUTONOMOUS), PUNE

(Affiliated to Savitribai Phule Pune University & Accredited by NAAC 'A' Grade)

2390-B, K.B. Hidayatullah Road, Azam Campus, Camp, Pune - 01 (Maharashtra, India)

National Conference on

GEOPOLITICAL DEVELOPMENTS AND GLOBAL ECONOMIC TRANSFORMATIONS : INDIA'S ADAPTATION AND RESILIENCE IN THE CHANGING WORLD

27th & 28th February 2023

Certificate

This is to certify that Prof./Dr./Mr./Ms./Mrs. Poonam Jadhav *from* Vishwakarma College of Arts, Commerce & Science *has participated & presented a research paper titled* Issues and Possibilities in marketing of agricultural commodities.

Vishwakarma College of Arts, Commerce & Science, Pune-48, PU/PN/ACS/ ... College Code No. 324

Ms. Anisa A. Khan HOD, Banking & Finance, Conference Co-ordinator	**Dr. Farzana V. Shaikh** HOD, Business Administration, Conference Co-ordinator	**Prof. M. G. Mulla** Dean, Faculty of Com. & Mgt. Conference Convenor	**Prof. Shaila Bootwala** Principal, Conference Chair

SARDAR PATEL INSTITUTE OF
ECONOMIC AND SOCIAL RESEARCH

anveṣak

A bi-annual journal

CERTIFICATE OF PUBLICATION

This is to certify that the paper entitled

ISSUES AND POSSIBILITIES IN MARKETING OF AGRICULTURAL COMMODITIES

Authored by

Poonam Jadhav, Dr. Shubhangi Auti

UGC

University Grants Commission

Approved Journal

vol. 53 No.01

in

Anvesak A bi-annual Journal

UGC Care Group - 1

ISSN: 0378-4568

January-June 2023

Impact Factor: 6.20

Editor
Anvesak
A bi-lingual Journal

Vishwakarma College of Arts Commerce & Science
ID No. PU/PN/ACS/275/2007
College Code No. 824
Pune-48