International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 5, Sept-Oct-2020, ISSN (Online): 2395-566X

Predicting and Managing Data using Data Analytics

S.Kaviarasan, E.Venkatesh, S.Shrinivasan, E.L.Sarath Pranev
Department of Computer Science and Engineering
Panimalar Institute of Technology, Chennai, India
arasan.kavi@gmail.com, venkatesh.e.281@gmail.com, shrinivasan2016@gmail.com, spranev@gmail.com

***Abstract* – Stock value forecast has been an inclining yet mystifying topic for quite a long period of time. These days developing of information become essential effort to store and maintain information. On the other hand due to availability of large number of data in all fields, it helps to manipulate the available data and makes that data to get useful by required way. Because of increased data volume information become significant job in all spots. By considering an under-used information source client can assume a job, for example, forecast of future investigation. Using corporate to analyze and build the predictor for future visit volume and for corporate predicts the stock price for product get falls up and down. By utilizing information sources we exhibiting the extravagance of the information source, shows how the information can be utilized to improve stock cost in corporate, finds intriguing patterns and conduct, gives stock cost to item finds a workable pace. The aftereffect of these examination are made utilizing KNN and mix calculation. This assists with understanding the upside of investigation of information and forecast report shows the information in visual diagram which makes the client to improve understanding.**

***Keywords* – Stock prediction, KNN, Forecast.**

I. INTRODUCTION

Stock price of the product is one of the difficult study to perform as the classification values are ranges in nature. There are many factors affecting the prediction behavior of share prices, such as physical factors, rational and irrational natures are also considered. These factors are affecting the volatility of the stock prices at greater levels and thus the price prediction becomes a major problem. Stock value expectation has been at research for quite a long time since it can return significant benefits. This review is about the specialty of building and utilizing likelihood models to take care of information driven issues.

We focus around idle variable models, which accepts that a complex data, which may be simple but unrecognized patterns. Latent variable models are analyzed in this study. The proposed work is comparable in similar to existing applications that are based on information discovered by gathering and analyzing volatility. The primary aim of our work is to classify the stock price according to age and product sales. Proposed work is an implementation of machine learning algorithm. Major study of classification systems give high accuracy with high handling time, though a few strategies give low precision even with enormous dataset. Along these lines, our work goes for high accuracy with immense dataset and less process time.The below figure, Figure 1, we represent the machine learning carried out over sales dataset. Nowadays, many efficient analysis techniques are available for affordable cost. These data analysis improves detection accuracy in stock price datasets.

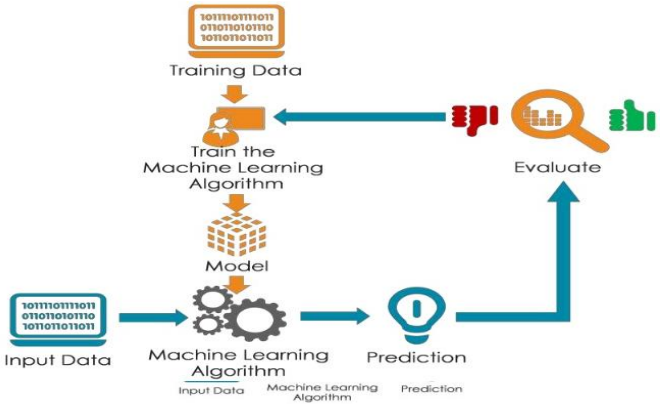

Fig.1. Machine learning Model.

Stock price details and its sales pattern analysis are useful for organizations to handle it in future sales and purchase. This study focuses on stock sales datasets by applying algorithm on our dataset. The remainder of this paper provides an overview of existing research handled by various authors in stock price prediction using machine learning approaches. Section 3 provides complete view of our implementation details. We end this study with a conclusion and future reference in section4.

II. EXISTING WORK

Numerous works has been proposed on stock price prediction and forecast systems exploiting machine

learning clustering, classification, information mining and learning. In this study, few of them is discussed with their concise proposition. One of the works done by Thissen et al. [1] in machine learning utilizing Support Vector Machine for stock price for time series prediction. The author discussed that the SVM is best on analyzing time series data than deep learning models such as neural networks.

Boser et.al. in [2] examined execution performance of margin classification algorithms namely polynomials, radial basis function, perceptrons. In their work they experiments optimal margin classifier. The classification technique on machine learning namely SVM and Random forest for stock price prediction is exploited by Sahaj Singh [3], they proposed forecast of price also through SVM, which derives the future price of the stock on particular day, for example next five days projection is discussed in this work.

Hiba et al [4] considered finding the best model for stock price prediction. They used two machine learning algorithms. Support Vector machine (SVM) and Random forest, which decides the quantity of fitting clusters Random Forest model is used for price prediction. The author through their experimental study proved that SVM prediction through best precision and recall values.

Osman et al, in [5] proposed Particle Swarm Optimization (PSO) model for stock price prediction. In their model, they proposed two correlation model one is PSO and another one is Least Square Support Vector Machine (LS-SVM). The author used thirteen dataset for comparative study and compared with Artificial Neural Networks and Levenberg-Marquardt (LM) algorithm on the considered dataset. Through experimental study, the author arrived that LS-SVM with PSO achieves good accuracy.

III. PROPOPOSED WORK

In our proposed work, we considered classification algorithm, K-nearest Neighbor (KNN) applied over considered Dataset. This dataset is analyzed using test set. Out of many available classification algorithms, KNN is preferred for its better accuracy on prediction part.
The following figure shows the methodology involved in our implementation.

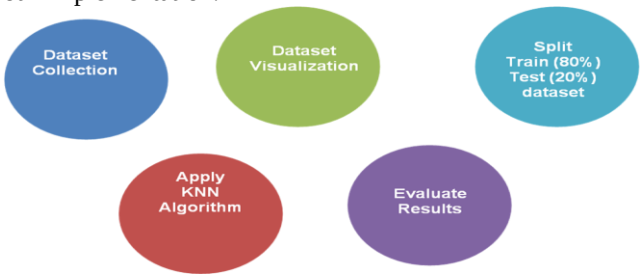

Fig.2. Overall Methodology of proposed work.

Implement machine learning KNN

1. The execution is carried out with following steps
2. Dataset collection
3. Dataset pre-processing
4. Split dataset as train and test set
5. Apply Machine learning and deep learning
6. Train the model
7. Give test set and predict values
8. Design a User Interface (UI) to give prediction

The below architecture, Figure 2, shows the overall flow of proposed work, which has the following advantages

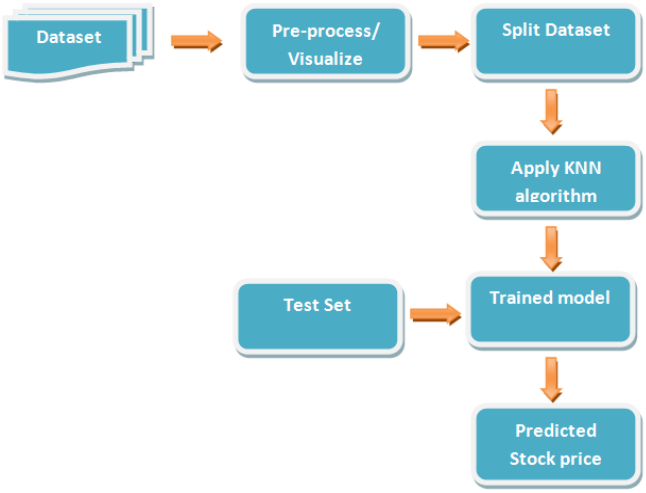

Fig.3. Overall Architecture of Proposed work.

The above figure, Fig 2 is the overall process handled in our proposed system. First we load the dataset, second it is pre-processed for any null values, third, pre-processed data is visualized plots. Then applied algorithm KNN and finally generating the output results with efficiency.
The above pre-processing steps make dataset ready to use for experiment.

1. Dataset Details

The below table, Table1 represents the characteristics of considered dataset. Missing attributes and Noisy attributes are not considered in our dataset.

Table -I: Dataset Characteristics

Data set	Attributes
ID	Customer ID
Gender	Customer Gender
Age	Customer Age
Salaryyu	Estimated Salary
Purchased	Number of products purchased
PID	Product ID

The below table, Table2 represents the feature attributes of our dataset along with their represented symbols.

Table -II: Dataset with sample values.

Id	Gender	Age	Estimated Salary	Purchased	Product Id
15624510	Male	19	19000	0	1
15810944	Male	35	20000	0	1
15668575	Female	26	43000	0	1
15603246	Female	27	57000	0	1
15804002	Male	19	76000	0	1
15728773	Male	27	58000	0	1
15598044	Female	27	84000	0	1
15694829	Female	32	150000	1	1
15600575	Male	25	33000	0	1
15727311	Female	35	65000	0	2
15570769	Female	26	80000	0	2
15606274	Female	26	52000	0	2
15746139	Male	20	86000	0	2
15704987	Male	32	18000	0	2
15628972	Male	18	82000	0	2

The dataset is ready to use as the feature values are already a numeric values.

III. IMPLEMENTATION METHODOLOGY

The dataset considered is split into train set and test set. We have considered 70% of data as training input for our machine learning algorithm and deep learning model to train the model. The remaining 30% of data is considered as test for result prediction.

1.K Nearest Neighbour (KNN) Algorithm

KNN algorithm, one of the supervised algorithm, which supports classification and regression is considered in our study. This algorithm use feature similarity for classification purpose.

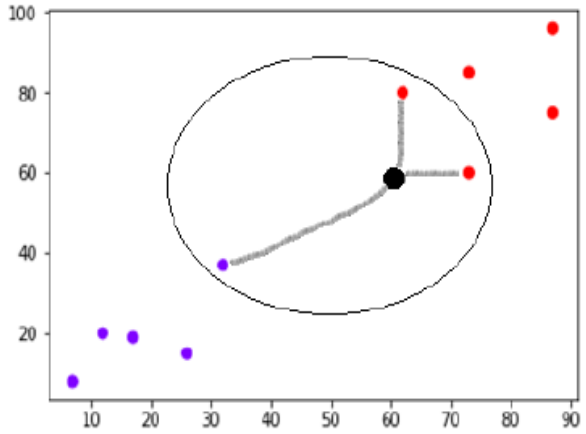

Fig.4. KNN Classification for 3 nearest neighbour.

The algorithm has following steps to follow
Step 1: Initialize and load the necessary data
Step 2: Initialize K as neighbors
Step 3:For each row in the data
Calculate the distance
Add (distance, index)
Step 4: Sort the collection by distances from the indices
Step 5: Select K entries from above collection
Step 6: Name the labels for the selected K values
Step 7: If regression, Consider mean of K values
Step 8: If classification, Consider mode of K values

IV. RESULTS AND DISCUSSION

The proposed work is implemented in ..... with mandatory libraries. The sales dataset is considered from this study. Machine learning algorithm KNN is used. We used these machine learning algorithm and indentified stock price and high sales product. The result shows that our model is more efficient. The following figure shows the sales of product for gender wise and some of the products shows used by male and some of the products are purchased by female in a large quantity are represented in this chart with sales values in Y-axis and Gender in X- Axis.

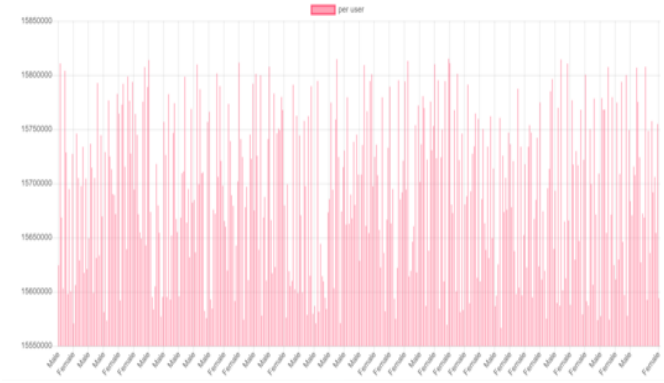

Fig.5 Gender wise sales data.

The following figure 6, shows the results metrics arrived from KNN algorithm which predicts the frequent sales items and plotted under. In this figure, X-axis represents product ID and Y-axis represents number of times sold.

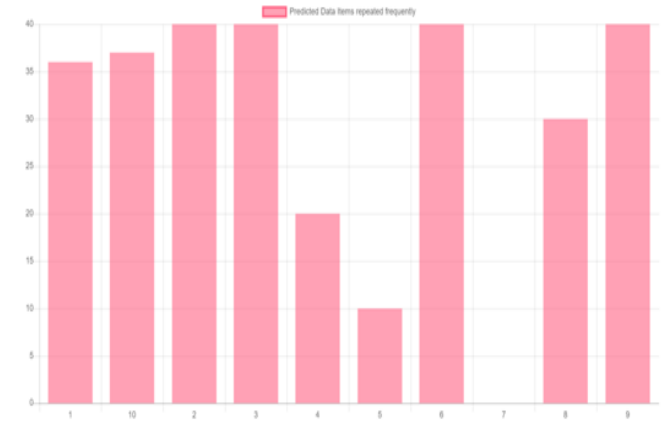

Fig. 6. Frequent Sold Items.

The following figure 7, shows the per user purchased items and numbers. The product id and number are mentioned here.

Fig.7. Per User purchase details .

The following figure 8, shows the purchase by age category. The category of age shown in multiples of 5 and the product ID they purchased are plotted here.

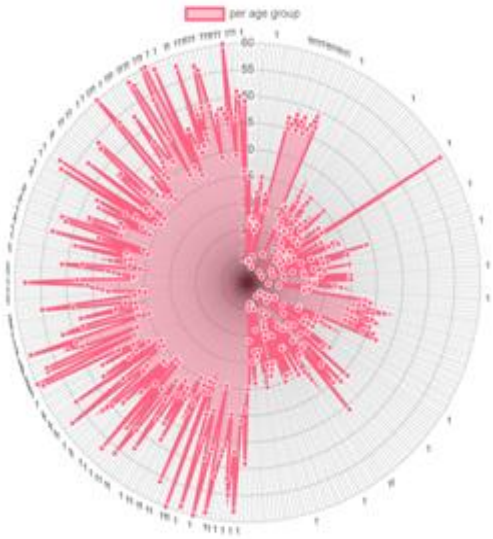

Fig.8. Purchase by Age category.

V. CONCLUSIONS

Stock data need to be processed to find out the pattern and extraction of data for analysis purposes, data mining and machine learning were used. In different sectors of stocks, these techniques were found useful including finding patterns, sales patterns, which age group using which product etc. In this work, we considered stock movement and purchase. We considered K- Nearest Neighbour algorithm (KNN), and evaluated its pattern as age wise and sales wise. Our work achieved high accuracy in KNN algorithm. Applying convolution model of deep neural network is our interest of further study on this research. Also this research can be extended to apply feature selection method before training the model.

REFERENCES

[1] Chang S V, Gan K S, On C K, et al. A review of stock market prediction with Artificial neural network (ANN). IEEE International Conference on Control System, Computing and Engineering. IEEE, 2014:477-482.
[2] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. I EEE Intelligent Systems and their applications , 13 (4), 18-28.
[3] Thissen, U., Van Brakel, R., De Weijer, A. P., Melssen, W. J., & Buydens, L. M. C. (2003). Using support vector machines for time series prediction. C hemometrics and intelligent laboratory systems, 6 9(1), 35-49.
[4] Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.
[5] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery , 2 (2), 121-167.
[6] Breiman, L. (1996). Bagging predictors. M achine learning, 2 4(2), 123-140.
[7] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news , 2 (3), 18-22.
[8] Schölkopf, B., & Smola, A. J. (2002). L earning with kernels: support vector machines, regularization, optimization, and beyond . MIT press.
[9] Du, W., & Zhan, Z. (2002, December). Building decision tree classifier on private data. In P roceedings of the IEEE international conference on Privacy, security and data mining-Volume 14 (pp. 1-8). Australian Computer Society, Inc..
[10] Weston, J., & Watkins, C. (1999, April). Support vector machines for multi-class pattern recognition. In ESANN (Vol. 99, pp. 219-224).

Sahaj Singh Maini, SCOPE, VIT, Vellore, India; Govinda.K, SCOPE, VIT, Vellore, India; Stock Market Prediction using Data Mining Techniques, Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017) IEEE Xplore Compliant - Part Number:CFP17M19-ART, ISBN:978-1