Sales Prediction with Data Mining Techniques

Arathy Ajith (c0790836)
Artificial Intelligence and Machine Learning
Lambton College of Toronto
Ontario, Canada
c0790836@mylambton.ca

Nidhi Parmar (c0796680)
Artificial Intelligence and Machine Learning
Lambton College of Toronto
Ontario, Canada
c0796680@mylambton.ca

Sachin Minhas (c0790285)
Artificial Intelligence and Machine Learning
Lambton College of Toronto
Ontario, Canada
c0790285@mylambton.ca

Elu Alias (c0786151)
Artificial Intelligence and Machine Learning
Lambton College of Toronto
Ontario, Canada
c0786151@mylambton.ca

Rincy Jose (c0790298)
Artificial Intelligence and Machine Learning
Lambton College of Toronto
Ontario, Canada
c0790298@mylambton.ca

Taranpreet Kaur (c0793771)
Artificial Intelligence and Machine Learning
Lambton College of Toronto
Ontario, Canada
c0793771@mylambton.ca

***Abstract*--- Data mining is the process of extracting hidden and predictive information from large databases; it is a powerful new technology that has the potential to assist businesses in focusing on the most important information in their data warehouses. Our proposed sales prediction system is used to identify the most common combinations of items present in the company. This will aid in marketing and sales, as well as the discovery of interesting cross-sells and related products.**

Keywords-- Sales Prediction, estimations, market, predictions

I. *INTRODUCTION*

.Sales forecasting is critical for most businesses, particularly car sales, real estate, and other day-to-day operations. On the basis of historical sales data, statistical methods such as regression or the autoregressive–moving-average (ARMA) are used to make predictions. These methods, however, are limited to a specific set of data. So many factors with complex interrelationships influence sales, and many of them are likely to be uncertain. We can identify potential models and development regularity from massive amounts of data using data mining. As a result, an increasing number of researchers are focusing on how to fully utilise data mining to process historical data and deal with trends in sales prediction.

II. *OBJECTIVE*

The overall goal of this project is to optimise sales prediction in order to improve organisational performance—more revenue, more profit, and higher customer satisfaction. In a broader view, the objectives include:

- To plan for the timely production of the right products in the right quantities and with the right quality.
- To plan for the timely procurement of high-quality raw materials, that can save a company a lot of money by keeping enough stock to meet sales forecasts.
- To forecast manpower planning, i.e. the number of casual/contractual labour required to ensure product availability.
- To calculate other expenses related to sales volumes, such as incentive costs, logistics costs, and stock maintenance.
- To plan sales promotion plans in order to increase sales in accordance with sales forecasts.
- To make logistical arrangements in advance to ensure that products are delivered to customers in good condition and on time.

III. *METHODOLOGY*

The methodology is the process/ML algorithm life cycle steps that we are going to follow in the project. Some steps are given below:

- To collect the data sets related to sales prediction and get the domain knowledge. Understand the project requirements.

- After that, handle the data cleaning, such as, check the missing values, handle the missing values, feature selection etc.
- In the next step, we can analyse the data by using some visualization tools. For example, to represent the data in the form graphs, charts, etc.
- Then, apply a machine learning algorithm, based on the data we can apply classification, clustering models etc.
- After applying models, we will do the evaluation of models and check the accuracy. If the accuracy is not good, we re-evaluate it and can try another model.
- Collect the data sets related to sales prediction from the various sources, such as Kaggle, UCI datasets, Github, etc. and get the domain knowledge. Understand the project requirements

In our research, we are planning to use algorithms such as Apriori, clustering, and classification to collect large amounts of highly volatile data from the servers of organisation. To group together various item sets, association rule mining techniques are proposed.

We intend to use the Walmart weekly sales dataset for the time being. We will test our methodologies on other similar datasets before finalising the dataset.

IV. *IMPLEMENTATION PLAN*

Sr.No	Tasks	Intended date of Completion	Assignee
1	Finalise project proposal and submit for clearance	16-03-2021	
2	collect data		
3	Process data and make preliminary interpretation		
4	Analyse data		
5	Developing suitable ML models		
6	Result analysis and comparison of best fit model		
7	Follow-up of implementation		

References

[1] http://ijcsit.com/docs/Volume%206/vol6issue03/ijcsit2015060368.pdf

[2] https://www.hindawi.com/journals/complexity/2021/6648009/

[3] https://ieeexplore.ieee.org/abstract/document/9342243