International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

# Precision Targeting: Social Media-Driven Profiling and Customer Segmentation for FMCG-Retail Industry

**CH .VENKATESWARLU [1], VENTURI RAMYA[2], GAVIREDDDY OBULA REDDY[3], SACHU VINAY[4], ANANTHASETTI AKHILA[5]**
#1Assistant Professor in Department of CSE in PBR VEC ,KAVALI.

#2#3#4#5 B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology & Science , Kavali.

**ABSTRACT_** Understanding consumer behavior and preferences is critical for effective marketing tactics in the retail industry of fast-moving consumer goods (FMCG). This initiative intends to use social media data to profile and segment clients in order to improve targeted marketing efforts. To find patterns in customer preferences, behavior, and demographics, we will analyze social media interactions such as posts, comments, and engagements using machine learning and natural language processing techniques.

The project will collect data from numerous social media platforms, extract features, and use clustering algorithms to divide customers into distinct groups based on their preferences, purchase behaviors, and engagement with FMCG products. The resulting consumer segments will provide useful information for personalised marketing efforts, product recommendations, and boosting overall customer happiness and loyalty in the FMCG retail industry.

## 1.INTRODUCTION

Stock-outs have long been a problem for retailers, distributors, and producers of fast-moving consumer goods (FMCG). Because FMCG products are not long-lasting, overstock situations are also undesirable. In order to boost demand for deteriorating goods, prices are frequently drastically lowered, resulting in a loss of potential revenue. In today's increasingly competitive retail industry, businesses aim to reduce, if not eliminate, stock-outs and overstocks. Overstocking results in high inventory costs and waste, while stock-outs have a direct impact on customer loyalty. Therefore, FMCG manufacturers, distributors, and retailers must coordinate their efforts in the processes of supply chain management in order to improve customer service quality and maximize efficiency. Regardless of whether sales are made in person or online, there is a vast amount of data related to demand. Retail is one of the first industries to use big data because of this. The properties of the FMCG data agree well with the referenced HACE theorem1. Scanners provide information on a wide range of aspects, including products, stores, marketplaces, and categories, to the typical retailer. These facts are used to create the demand hierarchy. There are many ways to divide the demand for FMCG.

## 2.LITERATURE SURVEY

[1] Rob J Hyndman, Earo Wang, Nikolay Laptev, It is becoming increasingly com- mon for organizations to collect very large amounts of data over time, and to need to detect unusual

or anomalous time series. For example, a large internet company has banks of mail servers that are monitored over time. Many measurements on server performance are collected every hour for each of thousands of servers. We wish to identify servers that are behaving unusually. We compute a vector of features on each time series, measuring characteristics of the series. The features may include lag correlation, strength of seasonality, spectral entropy, etc.

[2] Beatriz Pateiro-López, Alberto Rodríguez-Casal. The National Science Coun- cil (NSC) of Taiwan started the HAZ-Taiwan project in1998 to promote researches on seismic hazard analysis, structural damage assessment, and socio-economic loss estimation. The associated application software, "Taiwan Earthquake Loss Estima- tion System (TELES)", integrates various inventory data and analysis modules to fulfill three objectives. First, it helps to obtain reliable estimates of seismic hazards and losses soon after occurrence of large earthquakes. Second, it helps to simulate earthquake scenarios and to provide useful estimates for local governments or public services to propose their seismic disaster mitigation plans. Third, it helps to provide catastrophic risk management tools, such as proposing the seismic insurance policy for residential buildings.

[3] Toro-Gonzalez, Daniel McCluskey, Jill J. Mittelhammer, Ron Although mass- produced beers still represent the vast majority of U.S. beer sales, there has been a significant growth trend in the craft beer segment. This study analyzes the demandfor beer as a differentiated product and estimates own-price, cross-price, and income elasticities for beer by type: craft beer, mass-produced beer, and imported beer. We verify that beer is a normal good with a considerably inelastic demand and also find that the cross-price elasticity across types of beer is close to zero. The results suggest that there are effectively separate markets for beer by type.

**3.PROPOSED SYSTEM**

In Proposed System we used some machine learning algorithms random forest algo- rithm ,decision tree,knn algorithm

.

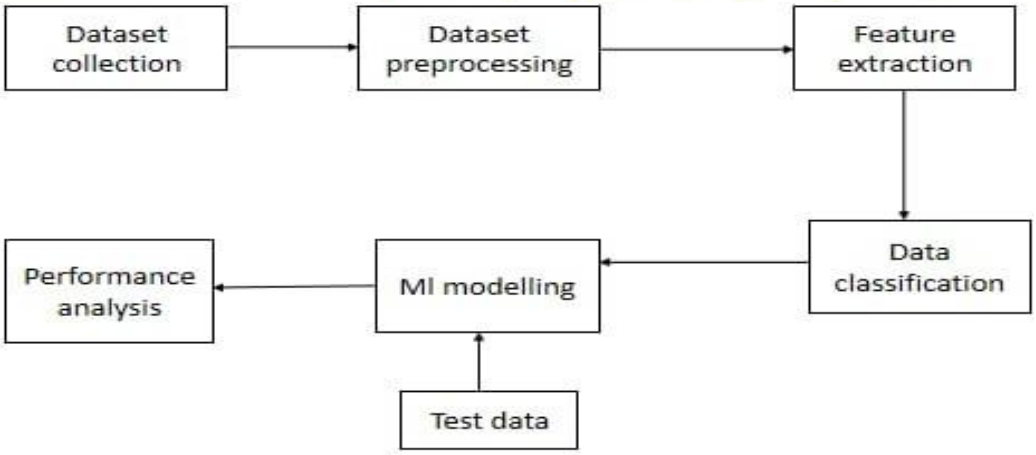



**Fig 1:Architecture**

The general structure of the predictive model is presented. At the start of this study, we only had the data set. After that, we forwarded the data set to the feature selection stage, where we received the most prominent and important features of our data for future use. After receiving the feature, we proceed to the following stage, which involves preparing the data set for the Random Forest algorithm's implementation. We will use a decision tree to train the data based on the best rules obtained.

**3.1 IMPLEMENTATION**

1.Support vector machine (SVM)

Support Vector Machines (SVM are machine learning algorithms that are used for classification and regression purposes. SVM are one of the powerful machine learning algorithms for classification, regression and outlier detection purposes. An SVM classifier builds a model that assigns new data points to one of the given cate- gories.Thus, it can be viewed as a non-probabilistic binary linear classifier.

2. Decision tree

The Decision Tree algorithm is one of the many algorithms that work on the concept of supervised learning. This algorithm can be used to solve both regression and classification-based use cases. It performs excellently when used in classification-based tasks in general and generates a tree-based structure on the dataset. As this algorithm mainly makes decisions based on certain factors that it considers to be important, it is quite relatable to the human way of thinking while making real-life decisions. The logic behind the decisions can also be easily understood due to the tree-like structure that the algorithm provides.

3. Random forest algorithm

The random forest algorithm in machine learning is a supervised learning algorithm. The foundation of the random forest algorithm is the idea of ensemble learning, which is mixing several classifiers to solve a challenging issue and enhance the model's performance. Random forest algorithm consists of multiple decision tree classifiers. First, each decision tree is trained individually. Then, the predictions from these trees are taken, and the random forest predicts the average of these re- sults.In the random forest

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

algorithm in machine learning, hyper parameters are usedto either speed up the model or improve its performance and predictive ability.

**3.2 MODULES**

**Module1 - Data Collection**

Data collection is a process in which information is gathered from many sources which is later used to develop the machine learning models. The data should be stored in a way that makes sense for problem. In this step the data set is converted into the understandable format which can be fed into machine learning models.

**Module2 - Data Pre-Processing**

Three common data pre-processing steps are:

Formatting: The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

Cleaning: Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed.

Sampling: There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

**Module3 - Feature Extraction**

Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed at- tributes, or features, are linear combinations of the original attributes. Finally, our models are trained using Classifier algorithm. We use classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.

**Steps to execute/run/implement the project**

**Step1 - Import Libraries**

In this project we are using the jupyter notebook for better output view when it comes to coding part first step is import the libraries example like numpy and pandas
, sklearns ,matplotlib,seaborn ect,.

**Step2 - Load Data**

To load the data we have library pandas( reading the data) and we get data set file

from the third party source and now you have to choose the file and upload and selects the data sets as requried by using df.head().

**Step3 - Transform and split**

Now transform the non-numeric data in the columns and selecting the specified attributes for new cleaned data set and split the data in to independent and dependent as x and y. Describe steps with title and mention steps in bullet points

**step4 - Feature Scaling**

Now setting a feature scaling and min-max scalar method scales the data set so that all the input and output features lie between 0 and 1 and splitting the data into train- ing(80

**step5 - Build the model**

The next step is build the model the model is randam forest algorithm, knn, decision tree and the model with no of decisions makining input and output. functions loss and accuracy and compile the model and next has to train the model.

## 4.RESULTS AND DISCUSSION

### 4.1 Efficiency of the Proposed System

As we have mentioned the earlier that it is better to search for an alternative approach with promising results and we are going to do that. So, in this study, we are going to compare four very commonly known machine learning algorithms and we are going to state the best among them using a PROFILING AND CUSTOMER SEGMENTA- TION FOR FMCG RETAIL INDUSTRY BASED ON SOCIAL MEDIA DATA As stated, the three machine learning algorithms are: Random Forest Decision Tree K- Near Neighbour (KNN) Why we are going with only 3 algorithms is that we don't want to overload and confuse the learning procedure by including more machine learning algorithms. So, we thought that three would be an optimal choice and the machine would work just fine.

### 4.2 Comparison of Existing and Proposed System Existing system:(Decision tree)

In the Existing system, The Decision Tree algorithm is one of the many algorithms that work on the concept of supervised learning. This algorithm can be used to solve both regression and classification-based use cases. It performs excellently when used in classification-based tasks in general and generates a tree-based structure on thedata set. As this algorithm mainly makes decisions based on certain factors that it considers to be important, it is quite relatable to the human way of thinking while making real-life decisions. But the accuracy of decision tree in existing system gives less accurate output that is less when compared to proposed system.

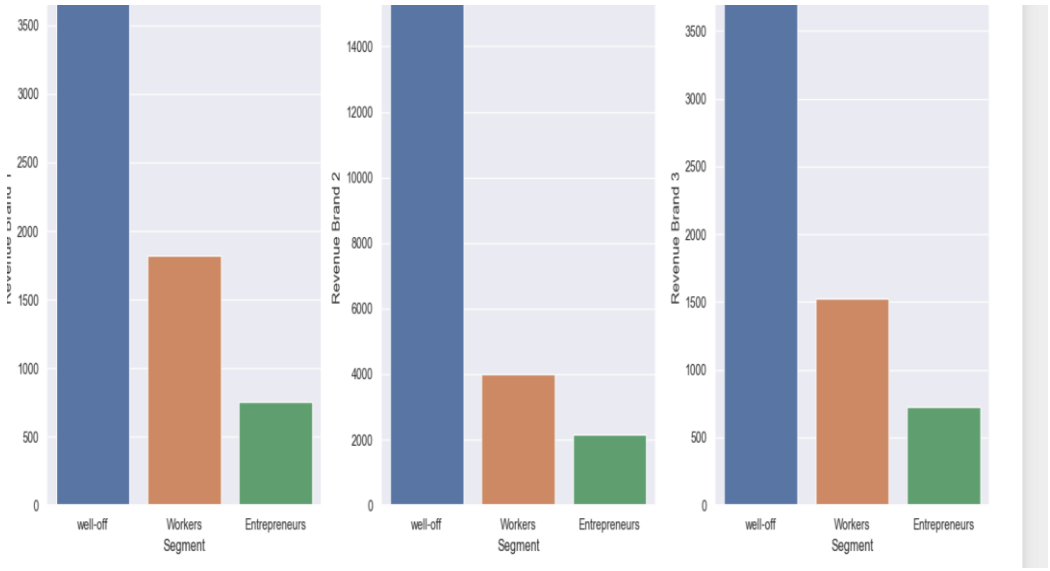
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

**Proposed system:(Random forest algorithm)**

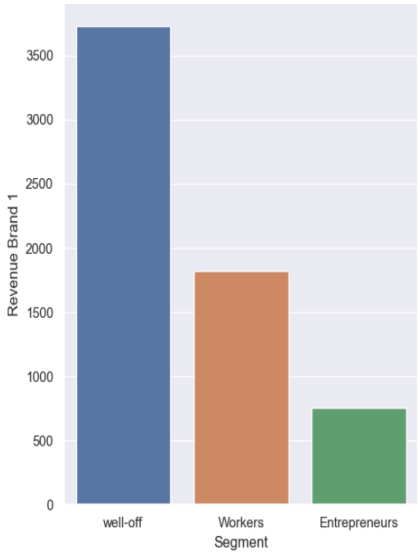
Random forest algorithm generates more trees when compared to the decision tree and other algorithms.The random forest algorithm in machine learning is a supervised learning algorithm. The foundation of the random forest algorithm is the idea of ensemble learning, which is mixing several classifiers to solve a challeng- ing issue and enhance the model's performance. Random forest algorithm consists of multiple decision tree classifiers. First, each decision tree is trained individually. Then, the predictions from these trees are taken, and the random forest predicts the average of these results. Proposed system is implemented using the Random forest algorithm so that the accuracy is more when compared to the existing system.

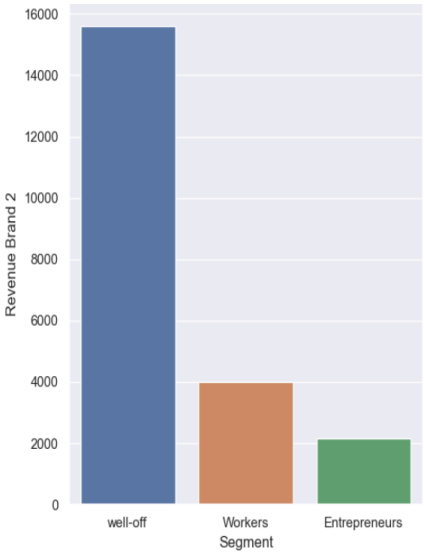
.

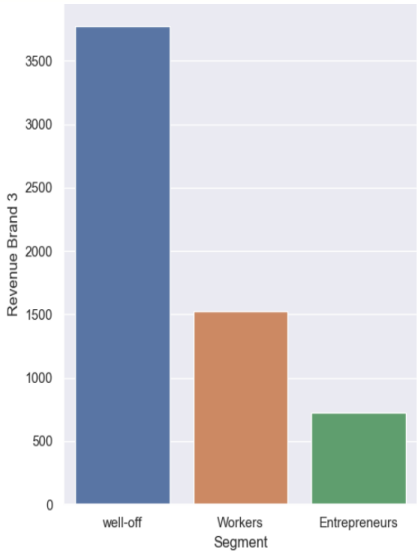
### 4.3 RESULTS

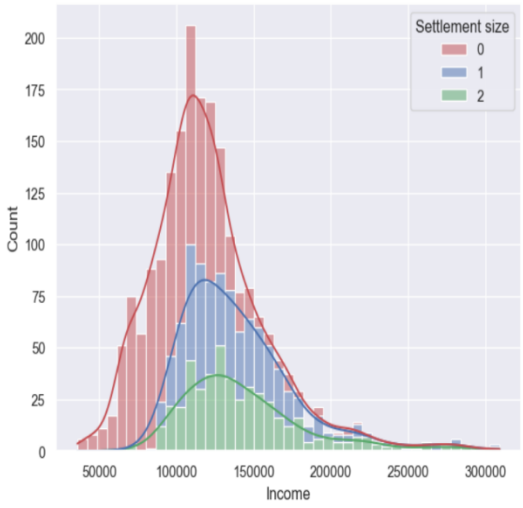


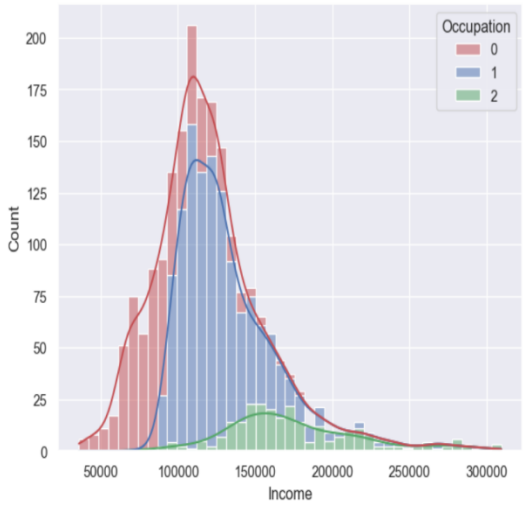
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

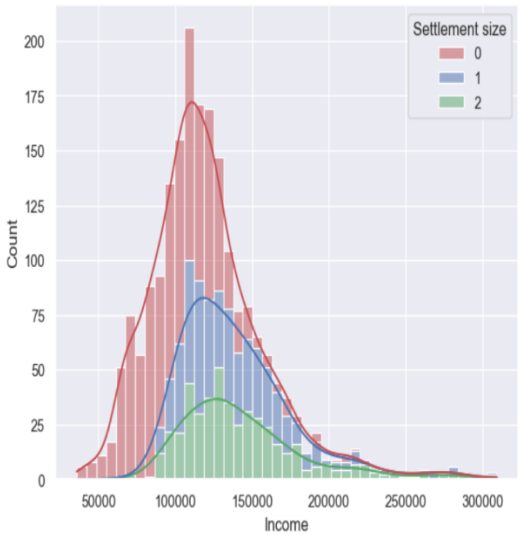


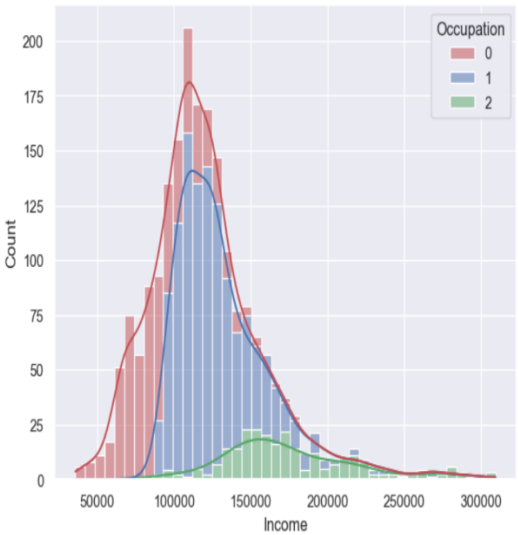



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362


## 5.CONCLUSION

Because of the hierarchical nature of demand in manufacturing supply chains, large amounts of time series data are usually present. This study recommends numerous exploratory analytics and visualization techniques for uncovering significant knowledge and information hidden in the data. When a single time series is of interest, we recommend superimposing exogenous features (such as pricing and promotion data) over the demand time series to graphically study the relationship between the factors and demand. We recommend utilizing the kite plot when working with a large number of time series. It is feasible to examine commonalities between distinct time series based on data availability.

## REFERENCES

[1] W. E. Wecker, "Predicting demand from sales data in the presence of stockouts," Management Science, vol. 24, no. 10, pp. 1043–1054, 2019.

[2] A. Chande, S. Dhekane, N. Hemachandra, and N. Rangaraj, "Perishable inven- tory management and dynamic pricing using RFID technology," Sadhana, vol. 30, no. 2-3, pp. 445– 462, 2020.

[3] D. Yang, G. S. W. Goh, S. Jiang, and A. N. S. Zhang, "Forecast UPC-level FMCG demand, Part II: Hierarchical reconciliation," in Big Data (Big Data), 2015 IEEE International Conference on, Oct 2018.

[4] T. Huang, R. Fildes, and D. Soopramanien, "The value of competitive informa- tion in forecasting FMCG retail product sales and the variable selection prob- lem," European Journal of Operational Research, vol. 237, no. 2, pp. 738–748, 2019.

[5] A. Jami and H. Mishra, "Downsizing and supersizing: How changes in product attributes influence consumer preferences," Journal of Behavioral Decision Mak- ing, vol. 27, no. 4, pp. 301–315, 2018.

[6] B. Pateiro-Lopez and A. Rodriguez-Casal., alphahull: Generalization of the Con- vex Hull of a Sample of Points in the Plane, 2019, r package version 2.0.

[7] R. J. Hyndman, "Computing and

graphing highest density regions," The Ameri- can Statistician, vol. 50, no. 2, pp. pp. 120–126, 2020.

[8] S. Thakur and T.-M. Rhyne, "Data Vases: 2d and 3d Plots for Visualizing Multi- ple Time Series," in Advances in Visual Computing, ser. Lecture Notes in Com- puter Science, G. Bebis, R. Boyle, B. Parvin, D. Koracin, Y. Kuno, J. Wang, R. Pajarola, P. Lindstrom, A. Hinkenjann, M. L. Encarnacao, C. T. Silva, and D. Coming, Eds. Springer Berlin Heidelberg, 2018, no. 5876, pp. 929–938.

**Author's Profiles**

**CH.VENKATESWARLU** working as Assistant Professor in Department of CSE, PBR Visvodaya Institute of Technology and Science KAVALI.

## Team Members

**Venturu. Ramya** B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology & Science, Kavali.

**Gavireddy. obula reddy** B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology & Science, Kavali.

**Sachu.Vinay** B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology & Science, Kavali.

**Ananthasetti .akhila** B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology & Science, Kavali.