Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

Customer Personality Prediction using the Ensemble Technique

Madarapu Soumica
Department of Computer Science
Maulana Azad National Institute of Technology
Bhopal, India

Bobbili Siva Rama Krishna
Department of Computer Science
Maulana Azad National Institute of technology
Bhopal, India

Chamarthi Somasekhar Varma
Department of Computer Science
Maulana Azad National Institute of Technology
Bhopal, India

***Abstract*—A comprehensive investigation of a company's ideal clients is known as a customer personality analysis. It aids a company's understanding of its consumers and makes it simpler to change goods to meet the demands, habits, and concerns of various sorts of customers. Customer personality analysis enables a company to adapt its product depending on the preferences of its target customers from various customer categories. Instead of paying money to promote a new product to every client in a firm's database, for example, a company may determine which customer group is most likely to buy the product and then market it to that segment solely. Hence, the main motive for the paper is to find the accuracy of the prediction of the personality of the customer who is shopping and improve the research out there using the ensemble technique.**

Keywords—Comparative analysis, machine learning,customer personality prediction, prediction t, ensemble model

I. INTRODUCTION

Doing business nowadays is not just waiting in a store and selling items. Day by day customers and their needs are also changing, and so the type of products according to them used to be changed.So, to overcome this problem we take help of Machine Learning algorithms to know more about customers and their needs. By using ML models which can predict a customer's needs depending upon their activities in a store or shopping mall and also on social media.For example, Instagram provides people to promote their business online and provide necessary tools. Tools may include paid promotion of their business, insights to monitor the post and stories activities such as how many people read their story or viewed their post.But now a question arises how this prediction thing works. There are algorithms like ANN, XGBoost and SVN in Machine Learning which works upon the datasets collected and based upon the algorithms make predictions about the customers' and likes and dislikes and displays them products according to it.Even prediction can also be made through the observation of customers by the security cameras in a store by observing how customers pretend on seeing a product and how time do they spend on seeing that particular product and also, they ask for the feedback from the customers. This helps us to gain more data about those products.Similar methods are used by almost everyone.

II. LITERATURE OVERVIEW

According a research ML in particularly two ways will be helpful to us in the prediction of the personality analysis.Firstly, by predicting values through the available datasets through which with are going to predict further predictions because it gives the data regarding the recent trends for our research.Secondly ML methods could allow us with the insights from personality psychology to be translated to practical applications in a more reliable way. There are other cross validation methods to provide us with more realistic estimation for how much accurate the model going to be.The usage of Personality prediction is going to be very useful in the near future as it will give us an idea about the customer expectation towards a product which will be beneficial especially for product-based companies. [1]Another study says ANN method will be helpful for the personality prediction from social media platforms like Facebook.Since lots of data of a person is already stored in a social media platform it is not even hard to predict anyone's emotion, likes and dislikes through that data. But that's the only limitation of this algorithm that should have enough data for the prediction. [2]Same as Facebook, Twitter also uses their data to predict people's behaviour by analysing their no. of followings and followers, no. of favourite tweets, their regular activities …etc. And on that basis, twitter suggests people to which page or person they should follow and gives its prediction about their likes and dislikes.And according to research there was a comparison between SVM method and XGBoost algorithm in which SVM gave 76.23% accuracy and XGBoost resultant 97.99% highest accuracy. [3]

III. METHODOLOGY

A. DATA PREPARATION

Data is an important component of the research and provides a future longitudinal advantage in analysis and finding a solution for an existing problem. Research tends to find new

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

discoveries and it isn't possible if data was not present. The dataset is taken from the customer future prediction dataset which has ~2240 entries and along with 29 features/attributes and some are ID,Year_Birth, Education, Marital_Status, Income, Kidhome, Teenhome etc and the rest can be seen in figure 1. The figure 2 gives a glimpse on how the dataset is looking and what data is there.

```
['ID',
 'Year_Birth',
 'Education',
 'Marital_Status',
 'Income',
 'Kidhome',
 'Teenhome',
 'Dt_Customer',
 'Recency',
 'MntWines',
 'MntFruits',
 'MntMeatProducts',
 'MntFishProducts',
 'MntSweetProducts',
 'MntGoldProds',
 'NumDealsPurchases',
 'NumWebPurchases',
 'NumCatalogPurchases',
 'NumStorePurchases',
 'NumWebVisitsMonth',
 'AcceptedCmp3',
 'AcceptedCmp4',
 'AcceptedCmp5',
 'AcceptedCmp1',
 'AcceptedCmp2',
 'Complain',
 'Z_CostContact',
 'Z_Revenue',
 'Response']
```

Figure 1: Attributes/features

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173

Figure 1: Dataset Overview

B. DATA PREPROCESSING

A dataset is made up of records, points, vectors, patterns, occurrences, instances, samples, observations, or entities. The key features of an item, such as the mass of a physical object or the time at which an event happened, are captured by a set of characteristics that constitute data objects. As a result, the first step towards improving dataset quality would be to check if there are any null values in the dataset which can cause a problem while giving the output and the author removed 24 null values from the income attribute. In figure 3 the dataset has been described and we can see that Z_constcontract and Z_revenue don't have any variance and hence can be removed from the dataset.

	count	mean	std	min	25%	50%	75%	max
ID	2240.0	5592.159821	3246.662198	0.0	2828.25	5458.5	8427.75	11191.0
Year_Birth	2240.0	1968.805804	11.984069	1893.0	1959.00	1970.0	1977.00	1996.0
Income	2216.0	52247.251354	25173.076661	1730.0	35303.00	51381.5	68522.00	666666.0
Kidhome	2240.0	0.444196	0.538398	0.0	0.00	0.0	1.00	2.0
Teenhome	2240.0	0.506250	0.544538	0.0	0.00	0.0	1.00	2.0
Recency	2240.0	49.109375	28.962453	0.0	24.00	49.0	74.00	99.0
MntWines	2240.0	303.935714	336.597393	0.0	23.75	173.5	504.25	1493.0
MntFruits	2240.0	26.302232	39.773434	0.0	1.00	8.0	33.00	199.0
MntMeatProducts	2240.0	166.950000	225.715373	0.0	16.00	67.0	232.00	1725.0
MntFishProducts	2240.0	37.525446	54.628979	0.0	3.00	12.0	50.00	259.0
MntSweetProducts	2240.0	27.062946	41.280498	0.0	1.00	8.0	33.00	263.0
MntGoldProds	2240.0	44.021875	52.167439	0.0	9.00	24.0	56.00	362.0
NumDealsPurchases	2240.0	2.325000	1.932238	0.0	1.00	2.0	3.00	15.0
NumWebPurchases	2240.0	4.084821	2.778714	0.0	2.00	4.0	6.00	27.0
NumCatalogPurchases	2240.0	2.662054	2.923101	0.0	0.00	2.0	4.00	28.0
NumStorePurchases	2240.0	5.790179	3.250958	0.0	3.00	5.0	8.00	13.0
NumWebVisitsMonth	2240.0	5.316518	2.426645	0.0	3.00	6.0	7.00	20.0
AcceptedCmp3	2240.0	0.072768	0.259813	0.0	0.00	0.0	0.00	1.0
AcceptedCmp4	2240.0	0.074554	0.262728	0.0	0.00	0.0	0.00	1.0
AcceptedCmp5	2240.0	0.072768	0.259813	0.0	0.00	0.0	0.00	1.0
AcceptedCmp1	2240.0	0.064286	0.245316	0.0	0.00	0.0	0.00	1.0
AcceptedCmp2	2240.0	0.013393	0.114976	0.0	0.00	0.0	0.00	1.0
Complain	2240.0	0.009375	0.096391	0.0	0.00	0.0	0.00	1.0
Z_CostContact	2240.0	3.000000	0.000000	3.0	3.00	3.0	3.00	3.0
Z_Revenue	2240.0	11.000000	0.000000	11.0	11.00	11.0	11.00	11.0
Response	2240.0	0.149107	0.356274	0.0	0.00	0.0	0.00	1.0

Figure 3: Description of the dataset

C. FEATURE SELECTION

Features are an important segment and the dataset which the authors have, have many features with them which help in making the model more accurate and better results. In figure 4 a correlation matrix is generated to know the relation between each feature or attributes and which feature is performing better with the help of this. And to double verify the data of the matrix the authors generated a heat map to see which features are mattering the most to the model's better output.

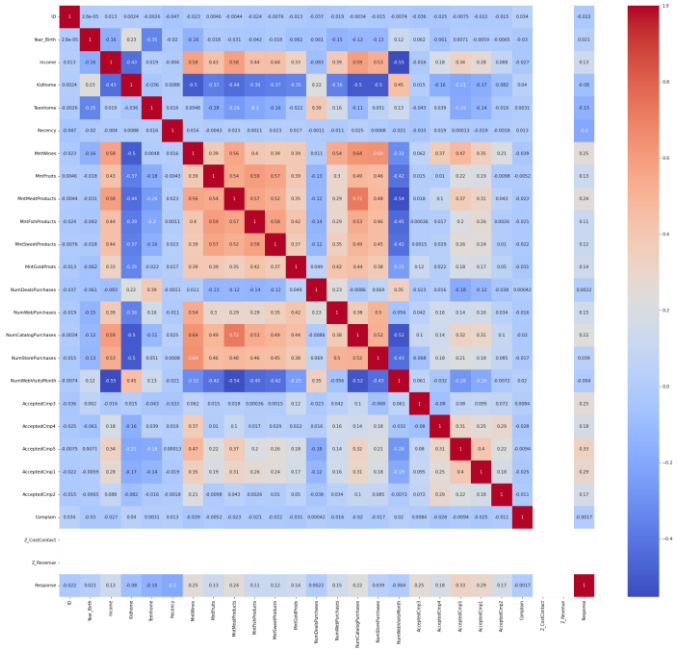

Figure 4: Correlation matrix

D. MODEL ARCHITECTURE

Ensemble model is also known as the combination of many models as it combines many such models into a single model and gets a better accuracy than all the models that are used in the process. In this paper the authors have created the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

ensemble model by combining support vector machine, naive bayes , logistic regression , knn and gradient boost. The accuracy of all the models alone was less than the ensemble model, Base estimators are the names given to these models. It's a way to get over the following technical difficulties in creating a single estimator. The features that this model provides is high variance, feature noise and bias.For a particular dataset, a single algorithm may not be able to provide the optimal forecast. Machine learning algorithms have constraints, and creating a high-accuracy model is difficult. If we build and combine multiple models, the overall accuracy could get boosted.Hence, this model was used and gave a good output in terms of accuracy.

IV. EXPERIMENTAL RESULTS

The author method was used in the algorithm presented and the stages involved to acquire the findings as we moved from data collection to preprocessing to feature selection to model creation and deployment. As the model used in the paper is an ensemble model the accuracy that is achieved after performing all these steps is ~99 percent which is very spectacular and better accuracy then other models out there.

```
Ensemble Model Accuracy: 0.9911904761904762
```

V. CONCLUSION

Customers are the primary income for the business and without them the business won't be able to survive in this competitive era. Hence, knowing what the customer wants based on his personality and likings makes the company's task easy to target him with that product and get him to buy that product. So, the model used by us gave an accuracy of ~99 percent, which would help many companies take this and apply it in their business.

VI. FUTURE SCOPE

Furthermore, we can use the following dataset to test out different regression models and neural networks and see how does it perform.Secondly, we can try this algorithm on a different dataset to know how does it perform and what problems it faces during the testing of the model.The research intended by the authors would help in the development of better and more productive and trustable prediction method.

VII. REFERENCES

[1] C. Stachl, G. Harari, S. Hilbert and R. Schoedel, "Personality Research and Assessment in the Era of Machine Learning".
[2] O. Ejimogu and S. Basaran, "A Neural Network Approach for Predicting Personality From Facebook Data"
[3] D. Suhartono, E. W. Andangsari, V. Ong and M. N. Suprayogi, "Personality Prediction Based on Twitter Information in Bahasa Indonesia".
[4] R. Bin Tareaf, P. Berger, P. Hennig and C. Meinel, "Personality Exploration System for Online Social Networks: Facebook Brands As a Use Case," 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2018, pp. 301-309, doi: 10.1109/WI.2018.00-76.
[5] K. Maheswari and P. P. A. Priya, "Predicting customer behavior in online shopping using SVM classifier," 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017, pp. 1-5, doi: 10.1109/ITCOSP.2017.8303085.
[6] P. Cong, G. Xu, J. Zhou, M. Chen, T. Wei and M. Qiu, "Personality- and Value-aware Scheduling of User Requests in Cloud for Profit Maximization," in IEEE Transactions on Cloud Computing, doi: 10.1109/TCC.2020.3000792.
[7] S. L. Christopher and H. A. Rahulnath, "Review authenticity verification using supervised learning and reviewer personality traits," 2016 International Conference on Emerging Technological Trends (ICETT), 2016, pp. 1-7, doi: 10.1109/ICETT.2016.7873647.
[8] R. D. Desai, "Sentiment Analysis of Twitter Data," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 114-117, doi: 10.1109/ICCONS.2018.8662942.
[9] A. Aslam, U. Qamar, R. A. Khan, P. Saqib, A. Ahmad and A. Qadeer, "Opinion Mining Using Live Twitter Data," 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2019, pp. 36-39, doi: 10.1109/CSE/EUC.2019.00016.
[10] K. Yan, X. Zhang, J. Tan and X. Fu, "Redefining QoS and customizing the power management policy to satisfy individual mobile users," 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), 2016, pp. 1-12, doi: 10.1109/MICRO.2016.7783756.
[11] S. Modi and M. H. Bohara, "Facial Emotion Recognition using Convolution Neural Network," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1339-1344, doi: 10.1109/ICICCS51141.2021.9432156
[12] C. Y. Yaakub, N. Sulaiman and C. W. Kim, "A study on personality identification using game based theory," 2010 2nd International Conference on Computer Technology and Development, 2010, pp. 732-734, doi: 10.1109/ICCTD.2010.5646417.
[13] R. S. Camati and F. Enembreck, "Text-Based Automatic Personality Recognition: a Projective Approach," 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 218-225, doi: 10.1109/SMC42975.2020.9282859..
[14] V. Shah and S. Modi, "Comparative Analysis of Psychometric Prediction System," 2021 Smart Technologies, Communication and Robotics (STCR), 2021, pp. 1-5, doi: 10.1109/STCR51658.2021.9588950.
[15] M. P. Aylett, A. Vinciarelli and M. Wester, "Speech Synthesis for the Generation of Artificial Personality," in IEEE Transactions on Affective Computing, vol. 11, no. 2, pp. 361-372, 1 April-June 2020, doi: 10.1109/TAFFC.2017.2763134.