Modeling Customer Purchase Behaviour

State-of-the-art Machine Learning tools and feature are utilized to Predict behavior of millennials and veterans on popular E-commerce sites

ME781 Project | Group-09

1) 213174001   **Hemant Rathore**
2) 22M1587   **Ashish Kumar Singh**
3) 20D100007 **Rushi Chavda**
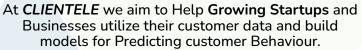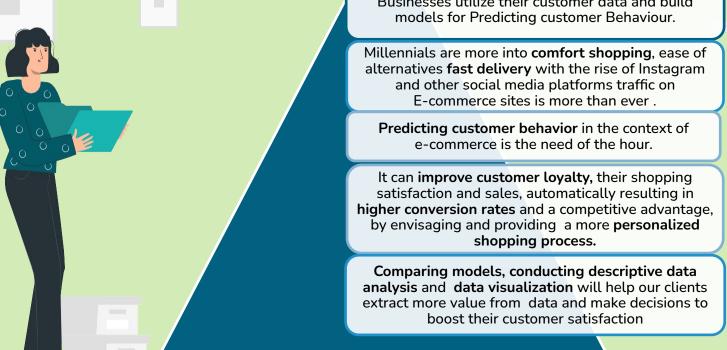4) 200020094 **Pragyesh Gupta**
5) 200040088 **Nidhi Chaudhary**

PROJECT OBJECTIVE

CLIENTELE

predict to progress

At *CLIENTELE* we aim to Help **Growing Startups** and Businesses utilize their customer data and build models for Predicting customer Behaviour.

Millennials are more into **comfort shopping**, ease of alternatives **fast delivery** with the rise of Instagram and other social media platforms traffic on E-commerce sites is more than ever .

Predicting customer behavior in the context of e-commerce is the need of the hour.

It can **improve customer loyalty,** their shopping satisfaction and sales, automatically resulting in **higher conversion rates** and a competitive advantage, by envisaging and providing a more **personalized shopping process.**

Comparing models, conducting descriptive data analysis and **data visualization** will help our clients extract more value from data and make decisions to boost their customer satisfaction

PROBLEM DEFINITION

CLIENTELE
predict to progress

☐ Millennials have become a major part of our customer base now and as the internet reach is widening into all generations, every person with a mobile is a potential customer for e-commerce sites. This shift makes **predicting customer behavior** all the more pertinent and gives you an edge in the competition.

☐ Once we help you **predict customer behavior** and customize their shopping process it will help **boost sales, increase customer satisfaction** and will certainly result in higher conversion rates and competitive **advantage.**

☐ At *Clientele* (http://www.clientele.predict), we utilize your customer data and build models for Predicting Customer Behaviour which will help Grow Your Startups and Businesses.

☐ We compare models to get further insight into the **performance differences in static customer data.** Conducting **descriptive data analysis visualization** will help our clients extract more value from data and make decisions to boost their customer satisfaction.

PROJECT DEFINITION

USP Its Protection

- Easy to use software
- Domain experts help
- High accuracy models
- Data protection and privacy
- Branding of USP
- AI Chatbot for support

Barrier to Entry and Existing Product/Services

- Companies not wanting to share data
- Companies building their own AI Teams
- Google Analytics
- Point Defiance Zoo
- Aquarium

Customer Requirements

- Accurate models
- User satisfaction
- Increasing revenue/campaign
- 24/7 Help and support

Business Case

- Target startups and small businesses
- Publish conclusions from publicly available data
- Subscription model like Bloomberg for companies
- Testimonials and references from satisfied clients

TECHNOLOGY LANDSCAPE ASSESSMENT

01 Patents

- Jivox Kairos™
- Predictive Intent Segments by Acxiom and AmEx

03 Open Libraries

- NumPy
- Pandas
- SciKit-Learn
- Matplotlib
- Seaborn
- Plotly
- Pydot
- XGBoost
- Unittest/pytest

02 Published Literature

1. Cirqueira D., Hofer M., Nedbal D., Helfert M., Bezbradica M. (2020). "Customer Purchase Behavior Prediction in E-commerce: A Conceptual Framework and Research Agenda." *Lecture Notes in Computer Science*, vol 11948. Springer, Cham.

2. Kumar, A., Kabra, G., Mussada, E.K. et al. "Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention." *Neural Comput & App* 877–890 (2019)

3. Dennis Koehn, Stefan Lessmann, Markus Schaal," Predicting online shopping behaviour from clickstream data using deep learning", *Expert Systems with Applications, Volume* 150,2020,113342.

4. Chen, Zhen-Yu, and Zhi-Ping Fan. "Distributed customer behavior prediction using multiplex data: a collaborative MK-SVM approach." *Knowledge-Based Systems* 35 (2012): 111-119.

PROJECT OUTCOMES

Approach and Outcomes :

- We took a **"customer_shoppers_intentions.csv"** and extracted independent features which affect the Odds of customer Purchase.
- We Did **Descriptive Data Analysis** on the data set to understand the data and trends into the data.
- **Data Processing** was done to convert data from categorical to numerical keeping in mind ordinal, nominal data types.
- The dataset is divided into **80:20 train-test ratio** and these algorithms are compared based on their accuracy on the test dataset.
- **Classification and Clustering** was used since outcome is binary.
- We Used Different Classifier to know which one is working best in aforementioned conditions.
- We Used all the Theories (Models) Taught In Class relevant to classification task.

Accuracy Summary:

1. Logistic Regression -- 88.6%
2. Random Forest -- 89.60%
3. XGBoost Classifier -- 89.05%
4. Gaussian Naive Bayes -- 61.71%
5. KNN classifier -- 87.71%
6. SVM Classification with PCA feature reduction technique -- 89.21%
7. SVM Classification with Feature Selection Dataframe -- 89.33%
8. Neural Network Classifier -- 89.61%

Discussions:

- These are the results we got for deployed models.
- We can see that the **Neural Network Classifier** is best among all the models.
- Random Forest, XGBoost Classifier and SVM classification are also giving good accuracy.
- Thus, we can choose the model, which have best trade-off between **Accuracy, robustness and speed.**