Ideation Phase
Empathize and Discover

Date	18-10-2023
Team ID	Team-591653
Project Name	Online Fraud Detection

Empathy Map Canvas:

In our empathy map centered on online fraud detection testing, our focus lies on a persona intricately linked with this subject matter. It's crucial to focus on personas intricately linked with the subject matter when it comes to online fraud detection testing. Online fraudulent transactions are a significant criminal violation that costs people and financial institutions billions of dollars every year. Financial institutions play a crucial role in detecting and preventing fraudulent acts, and machine learning algorithms provide a proactive way to prevent online transaction frauds with high accuracy. Online transaction fraud is a simple and easy target, especially with the increase in the number of online payment methods used in e-commerce and other online sites, raising the danger of online fraud. With the rise in fraud rates, machine learning approaches can be used to identify and evaluate fraud in online transactions. The primary goal of this project is to implement supervised machine learning models for fraud detection, analyzing prior transaction information, and classifying transactions into distinct groups based on the type of transaction. Various classifiers are trained independently, and models are assessed for correctness. The classifier with the highest rating score can then be picked as one of the best approaches for predicting fraud.

Empathy Map:

URL - Empathy Map: https://shorturl.at/GO158
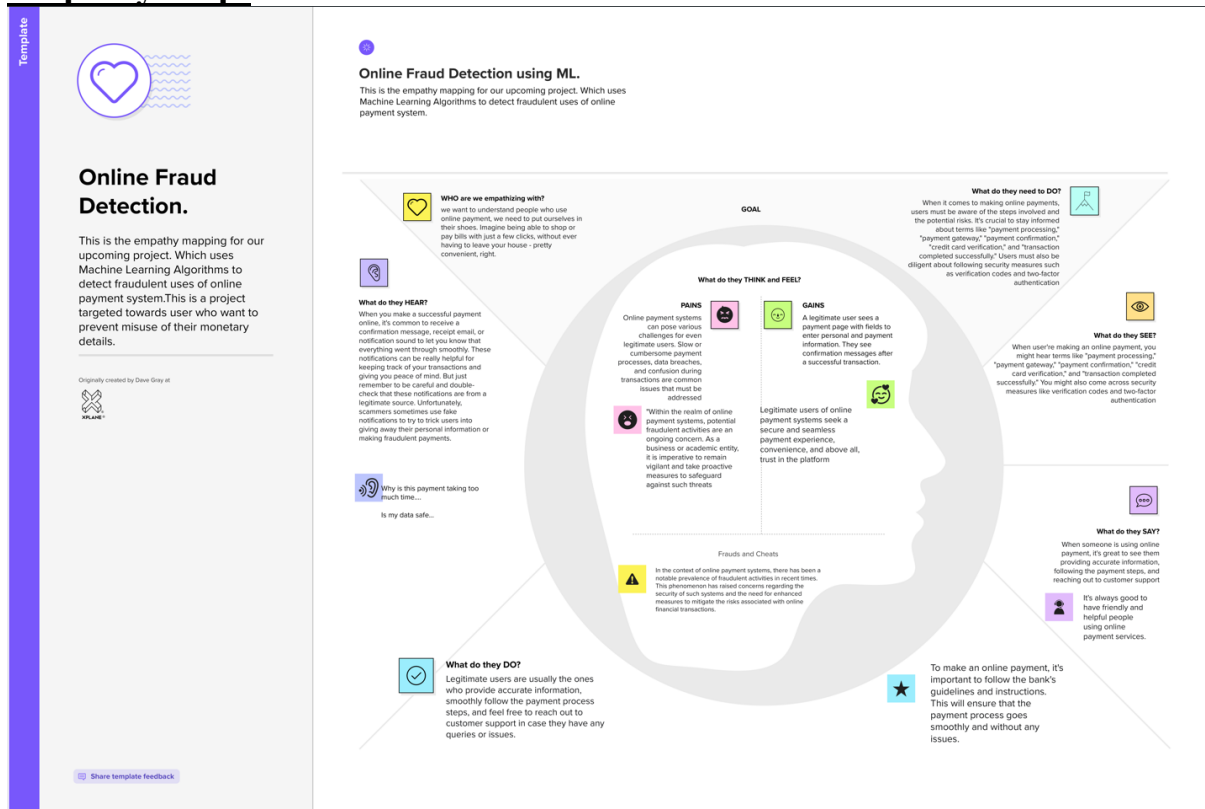
Ideation Phase
Brainstorm and Idea Prioritization

Date	18-10-2023

Team ID	Team-591653
Project Name	Online Fraud Detection

A group of people come up with a list of prospective ideas, proposals, or solutions during a brainstorming session for a project or problem.

Brainstorming:

After an extensive and thorough evaluation, we have decided and successfully identified and selected the final ideas for our project. Our selection criteria have been carefully honed to prioritize the ideas' significance and feasibility, and we are confident that this process has allowed us to choose ideas that not only hold immense potential but are also realistic and actionable. We will delve into the factors that influenced our decision-making, providing you with a comprehensive understanding of how we arrived at these definitive choices.

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

🕐 20 minutes

Random Forest
Random forest is an ensemble approach, because of its versatility and simplicity, it is one of the most often used algorithms. This model employs a large number of decision trees. Each of these decision trees separates a class of predictions, and the class with the most votes becomes our model's final output prediction. While growing trees in a random forest, rather than looking for the most significant characteristics for splitting, it seeks for the best features from a random selection of features for splitting the nodes. This results in a wide range of variety, which will provide us with a more accurate model. Because there is no association between the many models developed, the models provide ensemble forecasts that are more accurate than any of the individual projections. This is due to the fact that although certain trees may be incorrect.

K Nearest Neighbor
KNN is a non-parametric classification approach for solving classification and regression issues. KNN does not do any generalization, resulting in a relatively quick training procedure. Because of the lack of generalization, the KNN training phase is either small or retains all of the training data. The value k (number of nearest neighbors) is user defined.
K Nearest Neighbor algorithm is suitable for classification of fraud transactions, so by selecting the optimal nearest neighbor we can use K nearest neighbor to classify a transaction as legal or fraudulent.

Logistic Regression
Logistic regression is a classification procedure that is used to forecast the likelihood of a target variable. The target or dependent variable has a dichotomous character, which means there are only two potential classes. The representation for logistic regression is an equation. To anticipate an output value, input data are linearly mixed with coefficient values. The output value is represented as a binary value, which distinguishes it from linear regression

Support Vector Machine (SVM)
Support vector machine is a set of supervised learning methods used for classification, regression, and outlier detection. Different planes (hyperplanes) could be chosen, to separate the data points into two classes. Given a series of training examples, each labeled as belonging to one of two categories, an SVM training method constructs a model that assigns future instances to one of the two categories, resulting in a non-probabilistic binary linear classifier.

Artificial Neural Networks (ANNs)
These are a powerful tool in the realm of online fraud payment detection. ANNs are computational models inspired by the human brain's neural structure. They consist of interconnected nodes, organized in layers, which process and transform information. In the context of online fraud payment detection, ANNs are employed for several crucial tasks:

Decision Tree
A decision tree is a decision-making tool that employs a tree-like model of decisions and their potential outcomes, such as chance event outcomes, resource costs, and utility. It is one method of displaying an algorithm that consists solely of conditional control statements. Decision trees are a prominent method in machine learning and are often used in operations research, notably in decision analysis, to assist determine the approach most likely to achieve a goal.

Prioritization of ideas:

In order to decide which ideas should be implemented or pursued first, it is necessary to rank or evaluate them according to specified criteria, such as practicality, impact, cost, or strategic relevance.
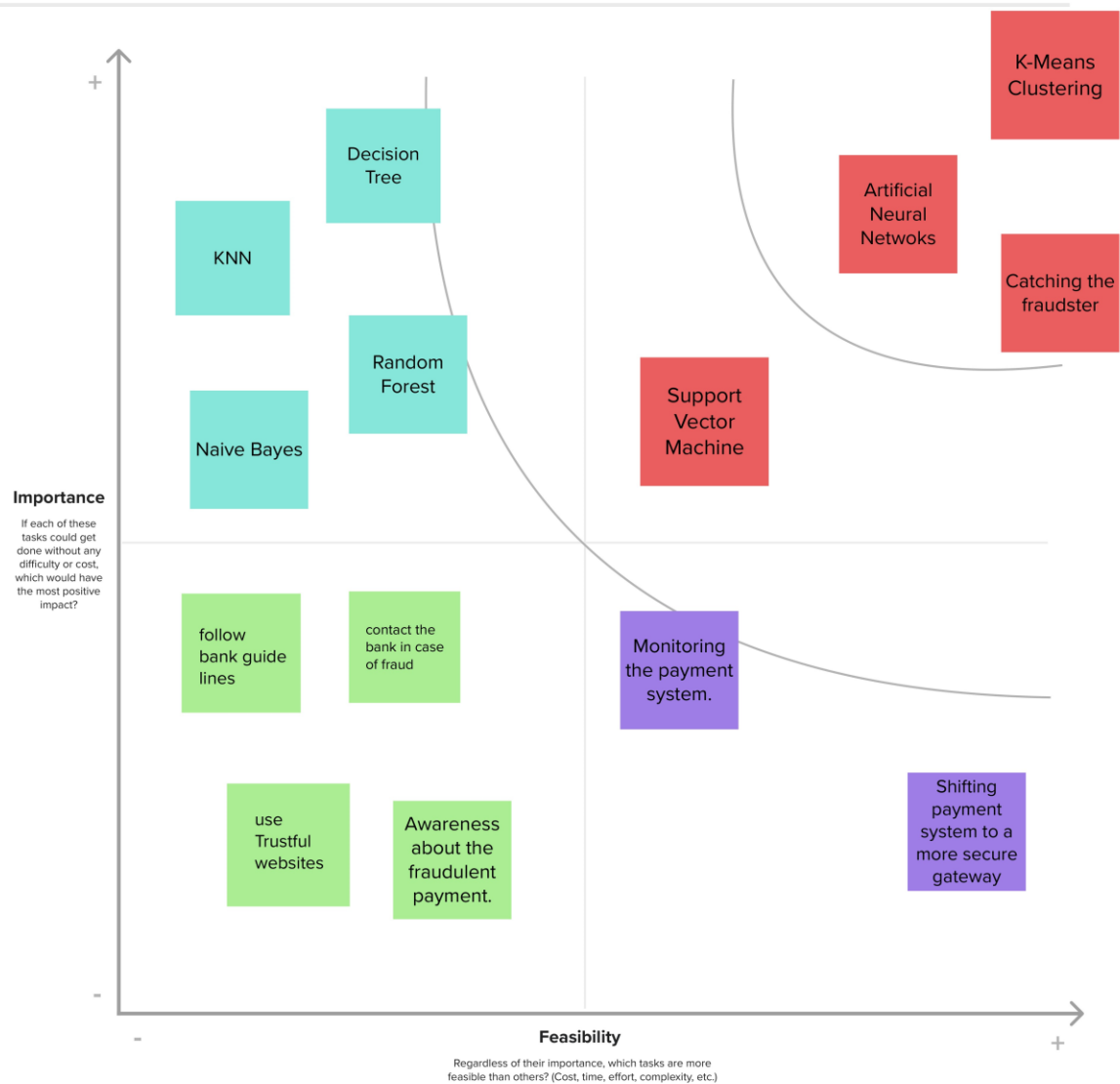
4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

🕐 20 minutes

+

K-Means Clustering

Decision Tree

Artificial Neural Netwoks

KNN

Catching the fraudster

Random Forest

Support Vector Machine

Naive Bayes

Importance

If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?

follow bank guide lines

contact the bank in case of fraud

Monitoring the payment system.

Shifting payment system to a more secure gateway

use Trustful websites

Awareness about the fraudulent payment.

-

-

Feasibility

Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

+

URL - Brainstorming and Idea Prioritization: https://shorturl.at/fsFX5