1. What problem your application solves? If it’s a new product, what is the size of the market it can address? If it A. I solution how much money can it save and what are the risks.

The application solves the Unsupervised problem by using clustering techniques. In any business it is important to know the target customers in order to increase sales and generate profit. Data such as Gender, age, annual income and spending score of the customers are taken into consideration during the analysis. Two models were used: K means and DBSCAN. The main aim is to use machine learning in business to increase the profit. Malls or supermarkets are often trying to increase their customer base in order to earn profits. A lot of them already have machine learning applied in order to analyze their customer base. They use their customer data and develop the ML models to target the appropriate audience. This application helps to focus on the target audience so that the mall or supermarket can focus on these customers by sending them email or SMS about their latest updates and offers.

We can save money by not triggering messages to the wrong audience. If a customer is not a potential buyer, there is no use of sending him/her a message about the promotions. This unnecessary wastage of money can be avoided and the mall can save a lot of money here.

As far as the risk is concerned, if the model fails to perform accurately, then the target customers may be neglected that may cause a loss to the mall. Other risk that is associated with the application is the security, if the application is breached and the customer data is leaked.

1. Explain your results: what was the performance of your method using metrics in class. Compare the results with other models’ example, Linear Regression vs Ridge Regression.

K- Means: The most famous clustering algorithm is K-means.

We get 5 clusters through K means approach. Each customer was placed in one of the five cluster. The k mean approach gives the following results.

#Cluster 1 -> clients with medium annual income and medium spending score

The customers belonging to this cluster are still not the prime targets of the mall. They have both average income and spending score. The mall will however try to increase their spending score.

#Cluster 2 -> clients with low annual income and low spending score

This is reasonable as customers who earn less will obviously try to spend less. These are the wise people who think before they spend. Hence, the mall will be least interested in the people belonging to this cluster.

#Cluster 3 -> clients with high annual income and low spending score [TARGET SET]

This is an interesting cluster. The customers belonging to this cluster are having high income and are still having a low spending score. They might not be happy with the services and the mall should try to add new facilities.

#Cluster 4 -> clients with high annual income and high spending score [TARGET SET]

The customers belonging to this cluster are considered to be the target audience. The mall will try to keep these customers happy as they are regular customers and are happy with the mall services.

#Cluster 5 -> clients with low annual income and high spending score

The customers belonging to this cluster love to shop heavily even though they have a low income. One of the reasons may be the services provided by the mall. The mall will not target these people with that enthusiasm but will try not to lose them as well.

Cluster size for K means:

Graphical user interface, application, table

Description automatically generated

Density-Based Spatial Clustering of Applications with Noise

It is based on the assumption that natural clusters are composed of densely located points.

Cluster size for DBSCAN:

Table

Description automatically generated

After comparing both the models, it is evident that DBSCAN failed to generate reasonable clusters. This is because it tries to find clusters based on the density of points. If one of our clusters is less dense than others DBSCAN will produce suboptimal results by not recognizing the least dense group as a cluster.

Concluding K means algorithms created better results.

1. What is the monetary value and Risks of your application after its performance? How much money can you save? Would you save money as you need fewer employees? You can estimate values if you can't find the data.

Monetary value of the application: $10,000

Risk of the application:

1. Security. As the application holds the data about the customers it is important to keep it secure. If the security is breached by the hacker, the customer data may be leaked.

Money saved from unnecessary triggering messages to wrong audience: $250/month

This application will not save any money with respect to the need of the fewer employees.

1. Other risks and benefits?

The benefits of mall customer data segmentation are as follows:

1. Personalization

This ensures that the mall provide exceptional customer experience to the customers especially the regular one.

1. Customer Retention

It is 16 times as costly to build a long-term business relationship with a new customer than simply to cultivate the loyalty of an existing customer.

1. Better ROI for marketing

Affirmations that right marketing messages are sent to the right people based on their life cycle stage.

1. Reveal new opportunities

Customer segmentation may reveal new trends about products and it may even give the first mover’s advantage in a product segment.