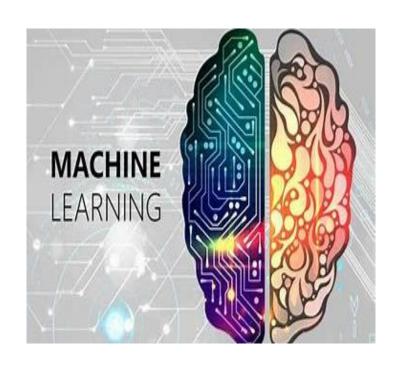# FINTECH BOOTCAMP Project - 3

Group members:

- Stephen Bridgens 1-3
- Peter Tran 4
- Roshan Paudyal 5-6
- Asha Devi 7-8

Presentation Date: 9 May 2023

EXECUTIVE SUMMARY

Objective: Develop a model to predict the success rate of customer purchases using machine learning.

Models Used: Neural Networks, K-Nearest Neighbours (KNN), and Random Forest Classifier.

Recommendation: Random Forest Classifier due to its highest Training Data Accuracy (99.09%) and Test Data Accuracy (89.38%).

Potential Application: Optimizing marketing strategies and resource allocation to increase sales and revenue.

CONCEPT

- Predicting customer purchase success rate using machine learning techniques.

- Leveraging variables related to user behaviour, website interactions, and administrative factors.

- **Goal**: Provide businesses with insights to optimize marketing strategies and resource allocation to maximize revenue.

DATA CLEAN-UP & PREPARATION

- **Dataset Source**: UC Irvine's Machine Learning Repository dataset.

- **Variables**: Page Values, Product Related Duration, Exit Rates, Bounce Rates, Product Related, Administrative Duration, Month, Administrative, Region, and Informational Duration.

- **Data Clean-up**: Handling missing values, outliers, and data inconsistencies.

- **Exploratory Data Analysis (EDA):** Understanding the distribution, relationships, and patterns within the dataset.

MODEL TRAINING AND EVALUATION

Models Used:

1. Neural Networks
2. K-Nearest Neighbours (KNN) - Unsupervised
3. Random Forest Classifier - Supervised

- Split dataset into training and testing sets and train each model using the training data.

- Evaluate model performance using accuracy metrics.

- Feature selection was attempted but resulted in lower prediction scores.

FINDINGS & CONCLUSION

Three classification models were compared for predicting the revenue: DNN, KNN, and Random Forest.

- All the models have accuracy above 85% - given using more features. High chances of purchase.

- The Random Forest model performed the best among the three in terms of accuracy, precision, recall, F1-score, and AUC-ROC on both training and test data.

- The DNN model had the highest accuracy on the training data, but its performance dropped on the test data.

- The KNN model also had a higher accuracy on the training data than on the test data.

Recommended model: Random Forest Classifier outperformed other models due to its high Training Data Accuracy (99.09%) and Test Data Accuracy (89.38%).

```
Classification:  Revenue ~ DNN

Training data:-
Accuracy: 0.8820965125709651
Accuracy_count: 8701
Precision: 0.8172888015717092
Recall: 0.27994616419919244
F1_score: 0.4170426065162907
AUC_ROC: 0.634422831443115

Test Data:-
Accuracy: 0.8633414436334145
Accuracy_count: 2129
Precision: 0.7814569536423841
Recall: 0.2796208530805687
F1_score: 0.41186736474694585
AUC_ROC: 0.6317380194952746
```

```
Classification:  Revenue ~ KNN

Training data:-
Accuracy: 0.8839213300892133
Accuracy_count: 8719
Precision: 0.7846410684474123
Recall: 0.3162853297442799
F1_score: 0.4508393285371702
AUC_ROC: 0.6504439300905692

Test Data:-
Accuracy: 0.85117599351176
Accuracy_count: 2099
Precision: 0.7165354330708661
Recall: 0.2156398104265403
F1_score: 0.33151183970856096
AUC_ROC: 0.5990136429823504
```

```
Classification:  Revenue ~ Random_Forest

Training data:-
Accuracy: 0.990977291159773
Accuracy_count: 9775
Precision: 0.9992852037169406
Recall: 0.9407806191117093
F1_score: 0.969150779896014
AUC_ROC: 0.9703306294412687

Test Data:-
Accuracy: 0.8937550689375506
Accuracy_count: 2204
Precision: 0.7515723270440252
Recall: 0.566350710900474
F1_score: 0.6459459459459461
AUC_ROC: 0.7638505022212742
```

POSTMORTEM

If no time constraint, we would like to achieve below points:

- Fine tune the model by adjusting the parameters and training parameters to optimize performance.

- Explore additional datasets or features that may enhance the model's predictive capabilities.

- Obtain more data to potentially increase the accuracy of scores.

Next Steps

- Utilize the developed model to help businesses optimize their marketing strategy and allocate resources effectively.

- Provide personalized recommendations to improve customer engagement and conversion rates.

- Continuously monitor and update the model to adapt to changing customer behaviors and market trends.

- Collaborate with businesses to implement the model and measure its impact on sales and revenue.

QUESTIONS?