

# INDIVIDUAL PROJECT - PREDICTING ONLINE PURCHASE INTENTION USING MACHINE LEARNING

Kaviya Balasubramaniam

# MOTIVATION & BACKGROUND

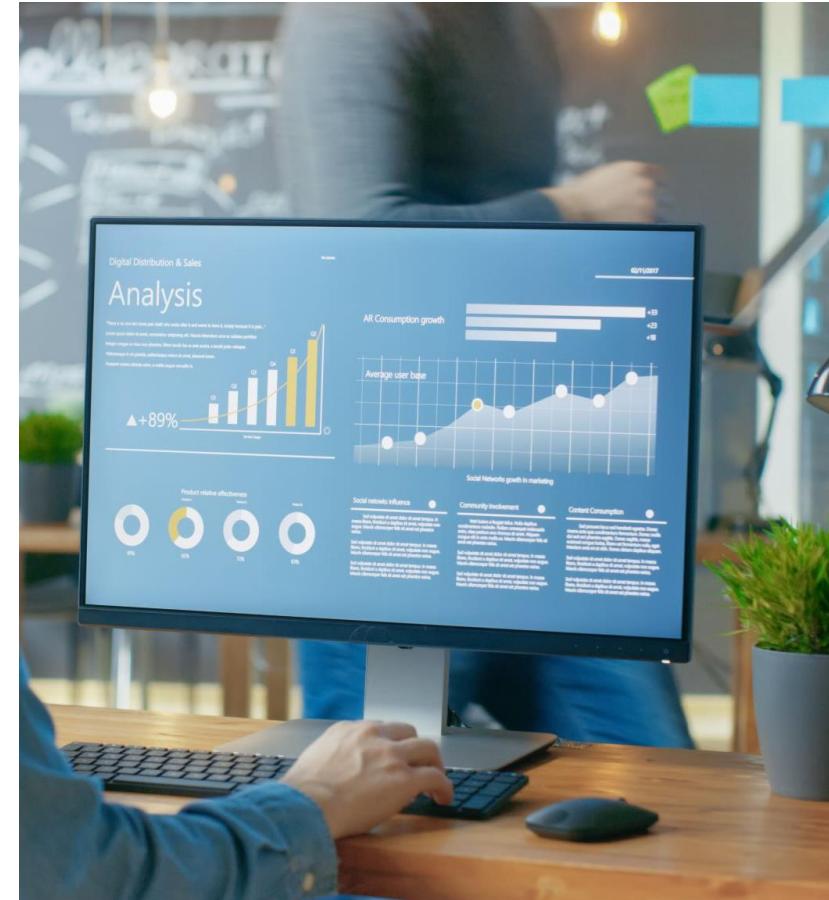
## Motivation and Background

Rapid growth of e-commerce generates large volumes of behavioural data

Predicting purchase intention is important for marketing and personalisation

Machine learning enables automated classification of user sessions

Purchase prediction is a well-studied problem in data science and business analytics





## RESEARCH QUESTION

- Primary question:
- Can machine learning models accurately predict whether an online shopping session results in a purchase?
- Sub-questions:
- Which model performs best under class imbalance?
- How do classical and modern models compare?

- **DATA OVERVIEW**

Online Shoppers Purchasing Intention dataset

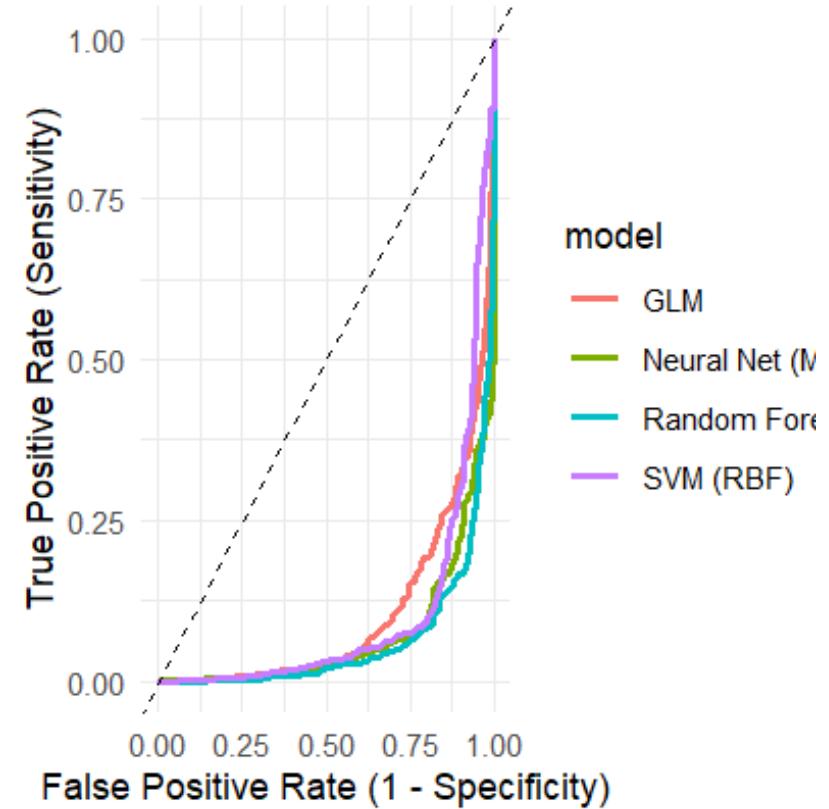
12,330 user sessions and 18 variables

Binary target variable: Revenue (TRUE / FALSE)

Behavioural, temporal, and technical predictors

Strong class imbalance ( $\approx 15.5\%$  positive cases)

ROC Curves (Test Set)



# METHODOLOGY

Methodology

Stratified 80/20 train-test split

Unified preprocessing:

Dummy encoding of categorical variables

Normalisation of numeric variables

Models applied:

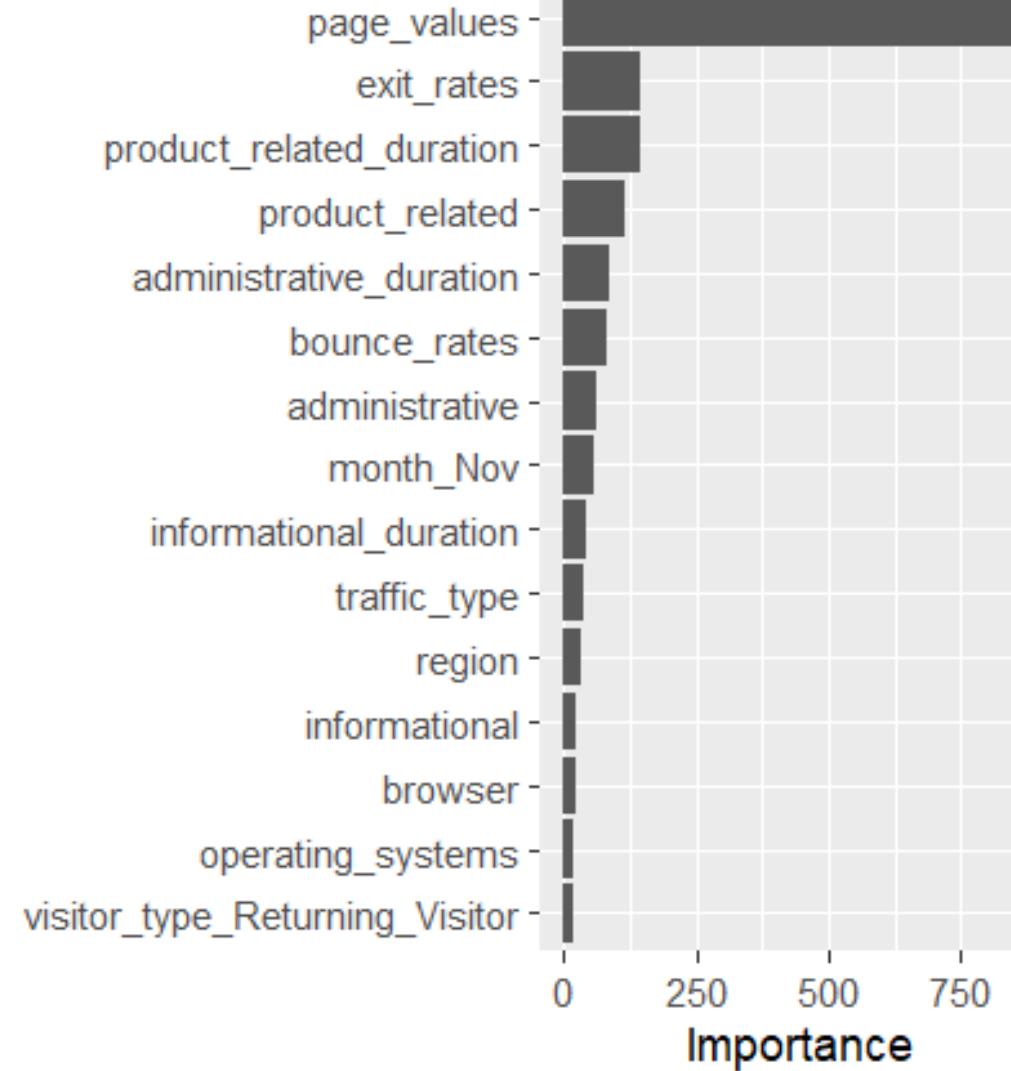
Logistic Regression

Random Forest

Support Vector Machine (RBF)

Neural Network (MLP)

Evaluation using ROC-AUC, PR-AUC, F1-score, sensitivity, and specificity



## RESULTS & CONCLUSIONS

- Results and Conclusions
- Random Forest achieved the strongest overall performance
- Neural Network performed competitively with non-linear patterns
- Logistic Regression provided an interpretable baseline
- SVM showed sensitivity to class imbalance
- Ensemble and neural models are best suited for purchase prediction

```
--- RF Confusion matrix ---
> print(rf_eval_kb726$confusion)
      Truth
Prediction FALSE TRUE
    FALSE  2013  140
    TRUE     72  242
> cat("\n--- RF Variable Importance Plot ---\n")
```