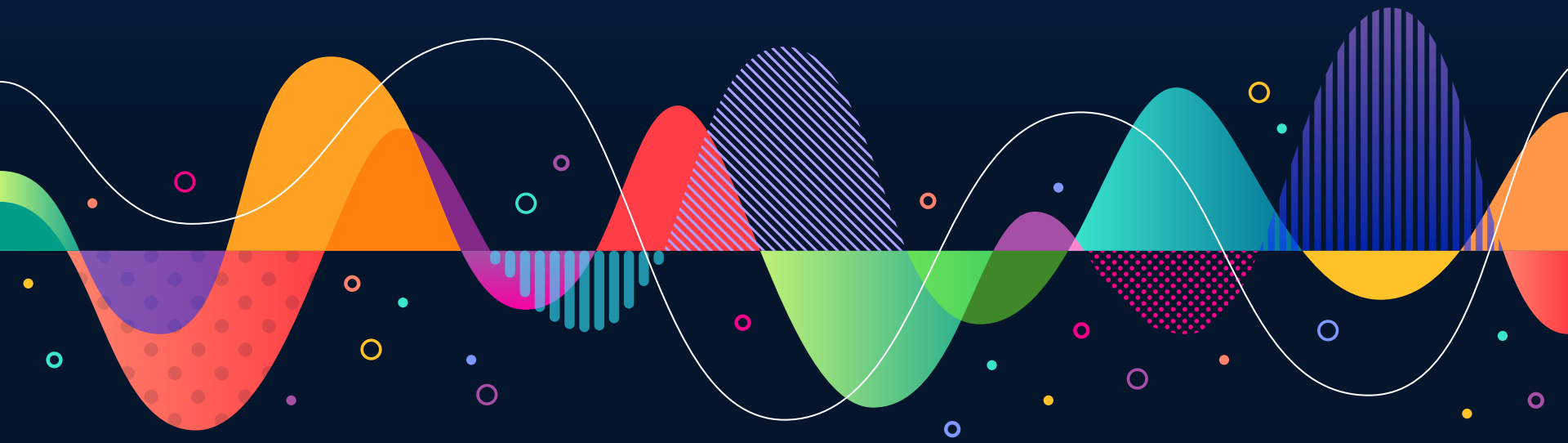
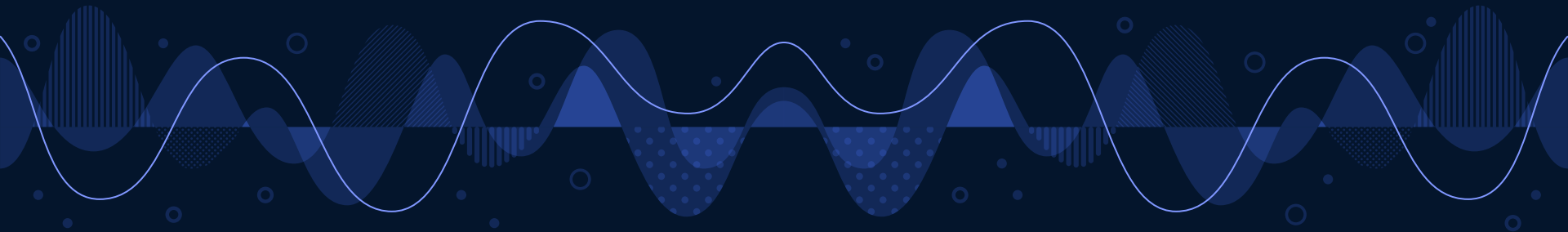
Model Accuracy vs Model Complexity

Pablo Valencia A01700912



Problem

- Complex models require more computational power
  - Simple models can outperform complex models to a certain extent
  - Data is to complex for a simple model
- Balance between accuracy and computational power



Problem

- Passenger satisfaction and loyalty depends on a huge number of factors including pre-flight, in-flight and post-flight services (Namukasa, 2013)
  - How can we determine which factors affect the most?
  - How can we predict customer satisfaction by changing certain factors

Dataset

- Airline Passenger Satisfaction from Kaggle
- 103904 surveys about customer satisfaction for training purposes
- 25976 surveys for testing purposes
- 26 features
- 2 classes (“neutral or dissatisfied” and “satisfied”)

Features

- Gender
- Customer Type
- Age
- Type of Travel
- Class
- Flight distance
- Inflight wifi service:
- Departure/Arrival time convenient
- Ease of Online booking
- Food and drink
- Online boarding
- Seat comfort
- Inflight entertainment
- On-board service
- Leg room service
- Baggage handling:
- Check-in service:

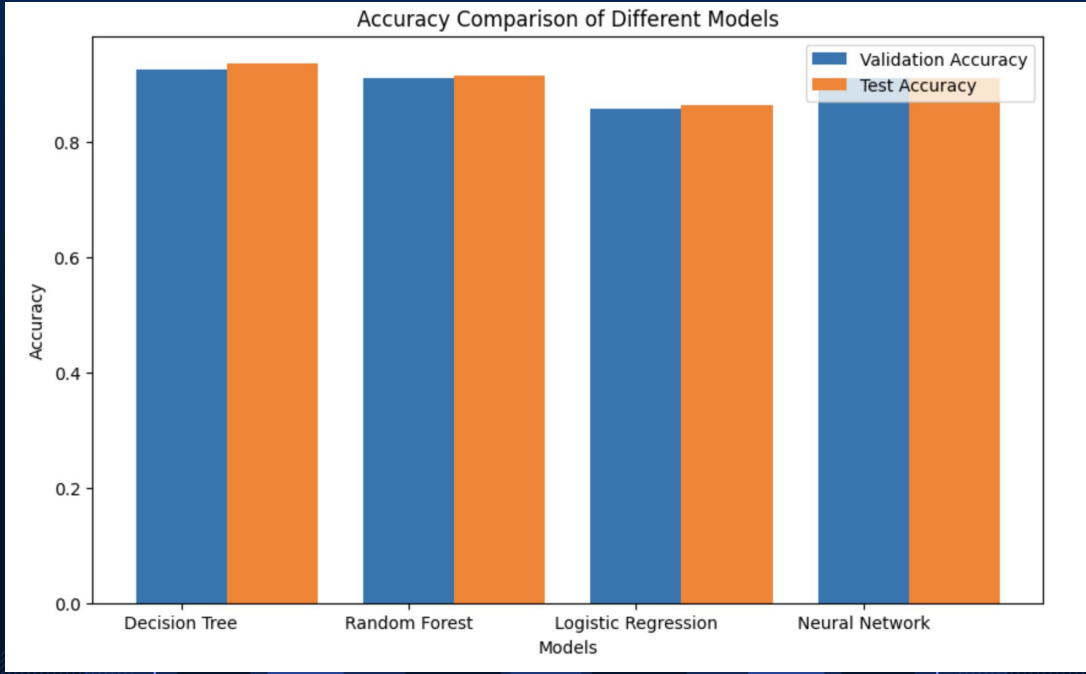
Preprocessing

1. Data exploration
    a. Check for missing values
        i. Drop rows
2. Convert from categorical to numerical
        i. Replace [“neutral or dissatisfied” and “satisfied”] with [0, 1]
        ii. One hot encoding of the following columns ’Gender’, ’Customer Type’, ’Type of Travel’, ’Class’

Trained Models

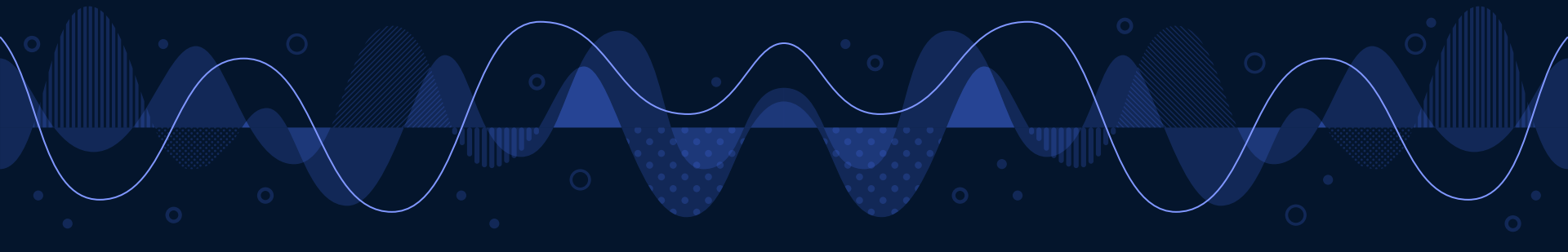
- Decision Tree
  - Max depth: 8
- Random Forest
  - Number of estimators: 500
  - Max leaf nodes: 16
- Logistic Regression
  - Trained for 500 epochs
- Neural Network
  - 20 epochs
  - 3 Layers (16, 16, 2)

Result


Accuracy Comparison of Different Models
Validation Accuracy
Test Accuracy
Accuracy
0.0
0.2
0.4
0.6
0.8
Decision Tree
Random Forest
Logistic Regression
Neural Network
Models

Conclusion

- Simple models can outperform complex models
  - This depends on the dataset
- This experiment requires way more analysis
  - CPU/GPU/RAM usage with each model
- Objective way to compare model complexity



References

- Namukasa, J. (2013), "The influence of airline service quality on passenger satisfaction and loyalty : The case of Uganda airline industry", The TQM Journal, Vol. 25 No. 5, pp. 520-532. https://doi.org/10.1108/TQM-11-2012-0092