SPOTIFY
Skip Prediction

By Faith Villarreal and Ricky Zapata

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Table of contents

01 Introduction

02 Data Analysis + Processing

03 Supervised Analysis

04 Conclusions

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Introduction: The Objective & Dataset

- **Objective:** Predict if users will skip or listen to a track.
- **Dataset:** "Spotify Sequential Skip Prediction Challenge" mini dataset from Spotify R&D; 170,000 user sessions.
- **Methodology:**
  - Utilizes 46 features from session and track data.
  - Focus on three 'skip' features ('skip_1', 'skip_2', 'skip_3') combined using OR logic for skip prediction.
  - Use labels (1) skipped and (0) not skipped for classification prediction/hypothesis and accuracy calculation.
- **Tools Used:**
  - pandas and numpy libraries for data manipulation.
  - sklearn for modeling
  - matplotlib for plotting graphs and data visualization

Mars Is a Cold Place
The 15th Planet

2:54   3:49

Data Analysis + Processing

- **Reorganizing Data:**
  - Combining 'skip_1', 'skip_2', and 'skip_3'
  - Linking sessions data to track data
  - Extracting year, month, and day of week from date feature
- **Ordinal Encoding:** *categorical features are ordered in nature so we can use Ordinal Encoding*
- **One-Hot Encoding:** *no order to categorical features so we can use One-Hot Encoding*
  - Used sklearn's OneHotEncoder
    i. *'context_type', 'hist_user_behavior_reason_start', 'hist_user_behavior_reason_end', and 'key' features*
- **Standardization of Data:**
  - Used sklearn's StandardScaler

Mars Is a Cold Place
The 15th Planet

2:54   3:49

Supervised Analysis: Logistic Regression

03

Feature Transformation
- **Polynomial. Degree of 2.**
- **Square Root**

L1 and L2 Regularization
- **C: [0.001, 0.01, 0.1, 1, 10]**

Best Accuracy: ~98%

Dataset	Square Root Transformation Degree	
	Original	**Transformed**
Train	0.9791592208720514	0.9791666666666666
Test	0.9791592208720514	0.9791666698776302

ROC-AUC Score: 0.98476
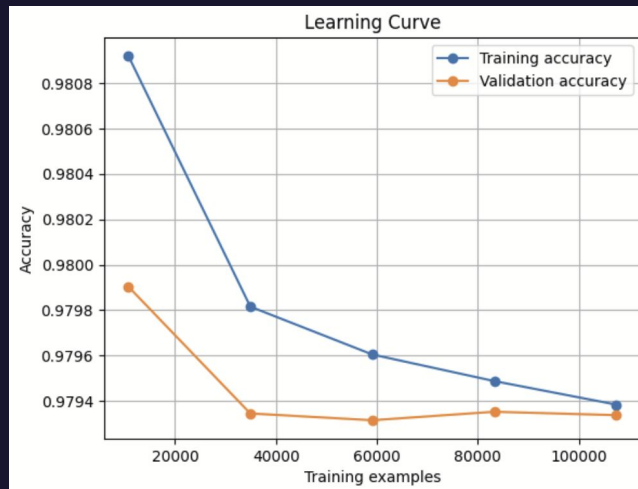
Supervised Analysis: SVM

03

Kernel Functions
- **Linear**
- **Polynomial. Degree of 2.**

Regularization
- **C: [100, 20, 5, 1 , 0.01, 0.001]**

Best Accuracy: 98%

Linear. C = 1.

Mars Is a Cold Place
The 15th Planet

2:54

3:49

03

Supervised Analysis: Neural Network

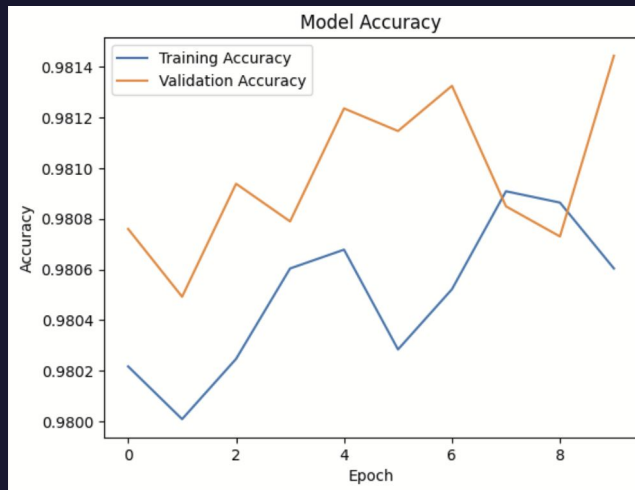## Activation Functions
- ReLu
- Sigmoid

Regularization
- Ridge: 0.01, 0.001
- Lasso: 0.01, 0.001

Layers
- [64, 32, 16, 1]
- [64, 32, 1]

Best Accuracy: 98.14%

4 Layers. No regularization.

Mars Is a Cold Place
The 15th Planet

2:54   3:49

04

Conclusions

- Neural Networks performed best.
- Regularization did not help (much)
- Generalized Question = Generalized Answer
- Massive Dataset

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Thank You!

Mars Is a Cold Place
The 15th Planet

2:54

3:49