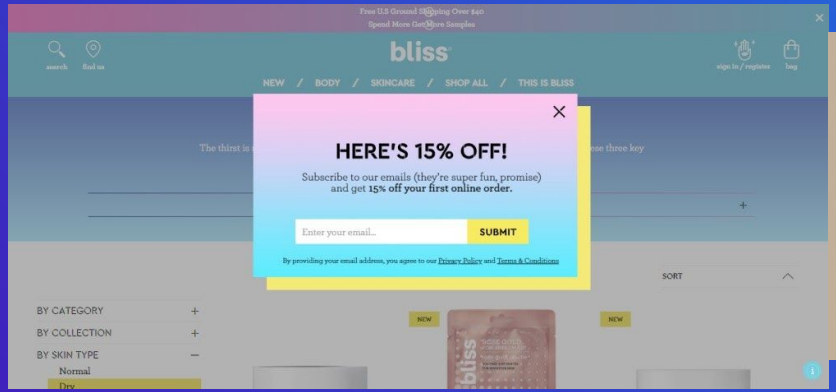
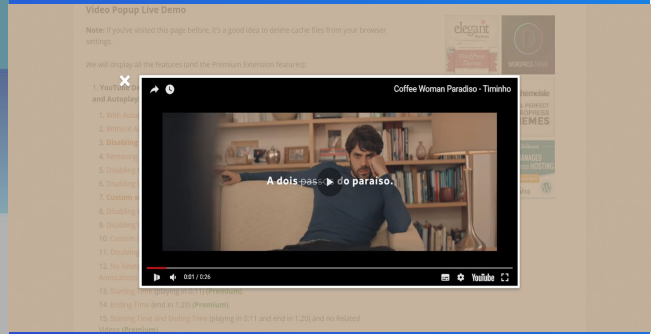
Real-Time Prediction of Online Shoppers' Purchasing Intents Using K-Nearest Neighbor Classification

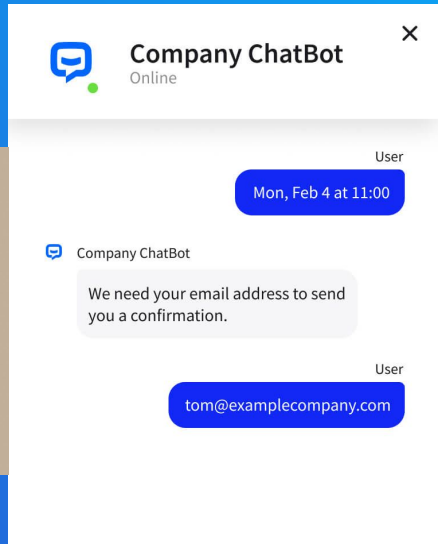
Pranav Tonpe

Problem

- **3 out of 4** online users in the US, UK, and several European countries believe that online ads are irrelevant (often times, this pop-up media is a result of certain website AI algorithms)
- **4 out of 5** shoppers have left a webpage because of bothersome pop-up ads and auto-play videos prompting them to buy a product
- **70%** of marketers are failing to target customers with behavioural data due to a lack of understanding between users and e-retailers

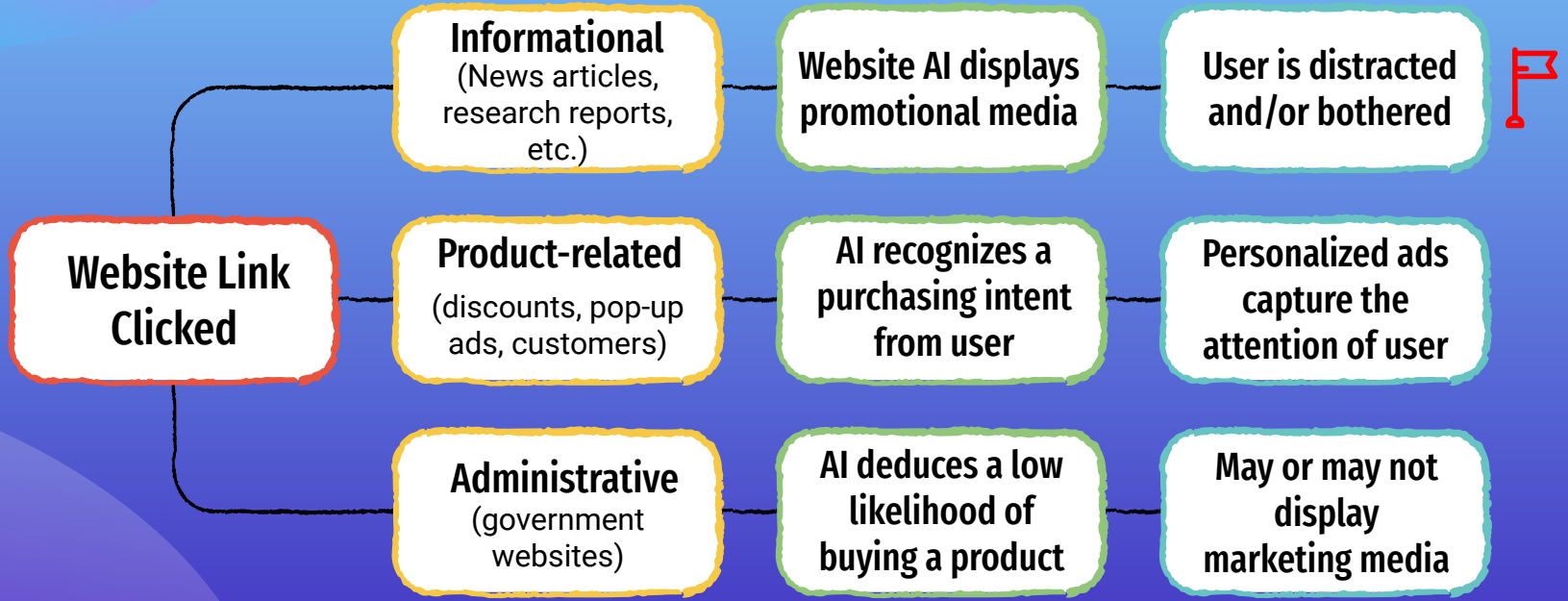






Situational Overview

Anonymous user surfing the Internet



Methodology

01

Data Processing

Load data into readable spreadsheet and identify key columns (use CLAs)

02

Feature Initialization

Convert dataset values (bounce rate, exit rate, duration, etc.) into float and int values

03

Data fitting

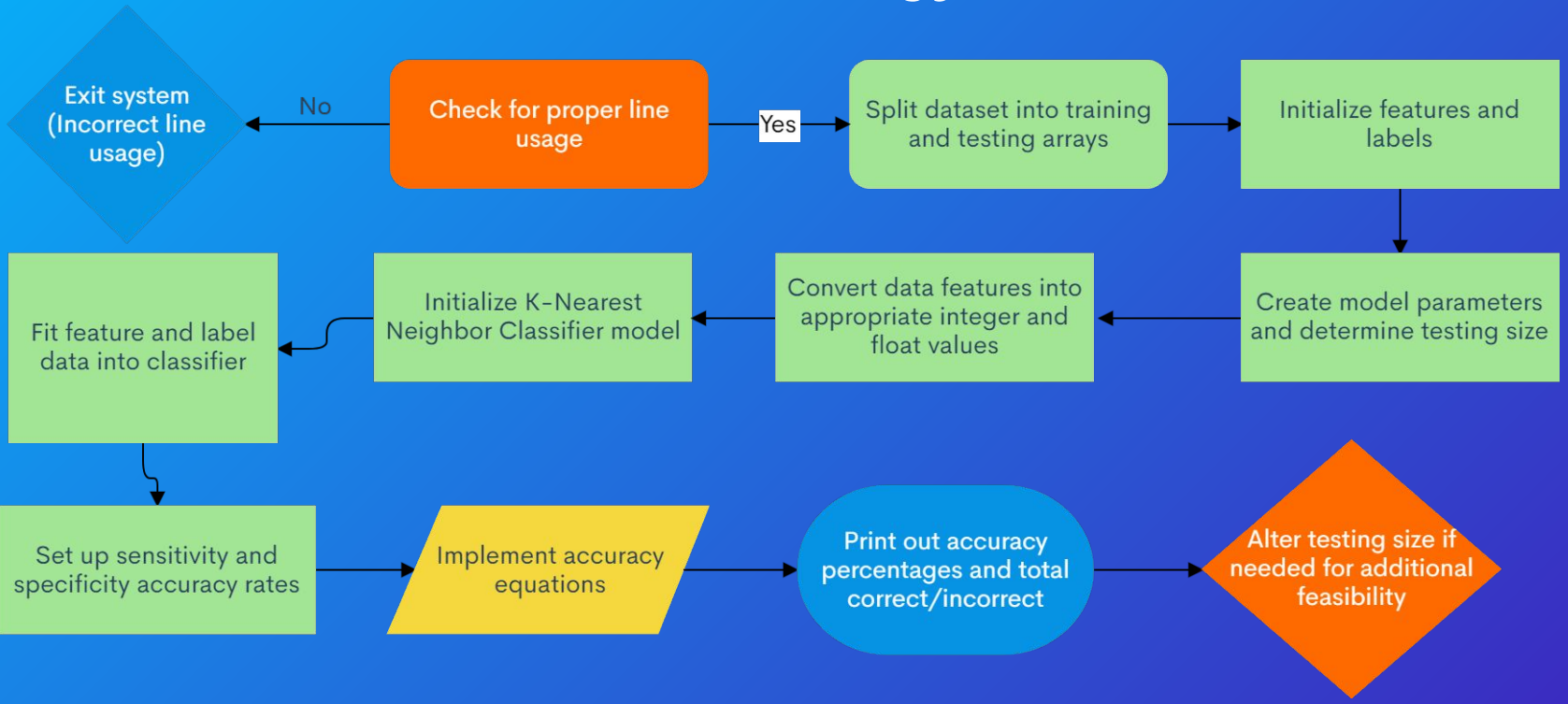
Create K-nearest neighbor model and analyze outcomes to make future predictions

04

Accuracy Metrics

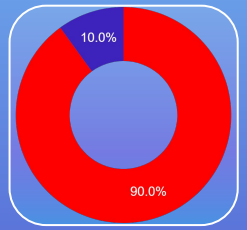
Initialize specificity and sensitivity rates and implement their equations

Methodology



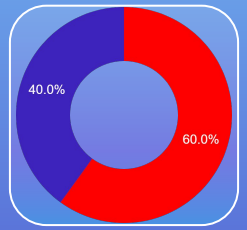
Ideal Situation

Results



89.4% TP
43.5% TN

Average accuracy in the 90-10 train-test split dataset

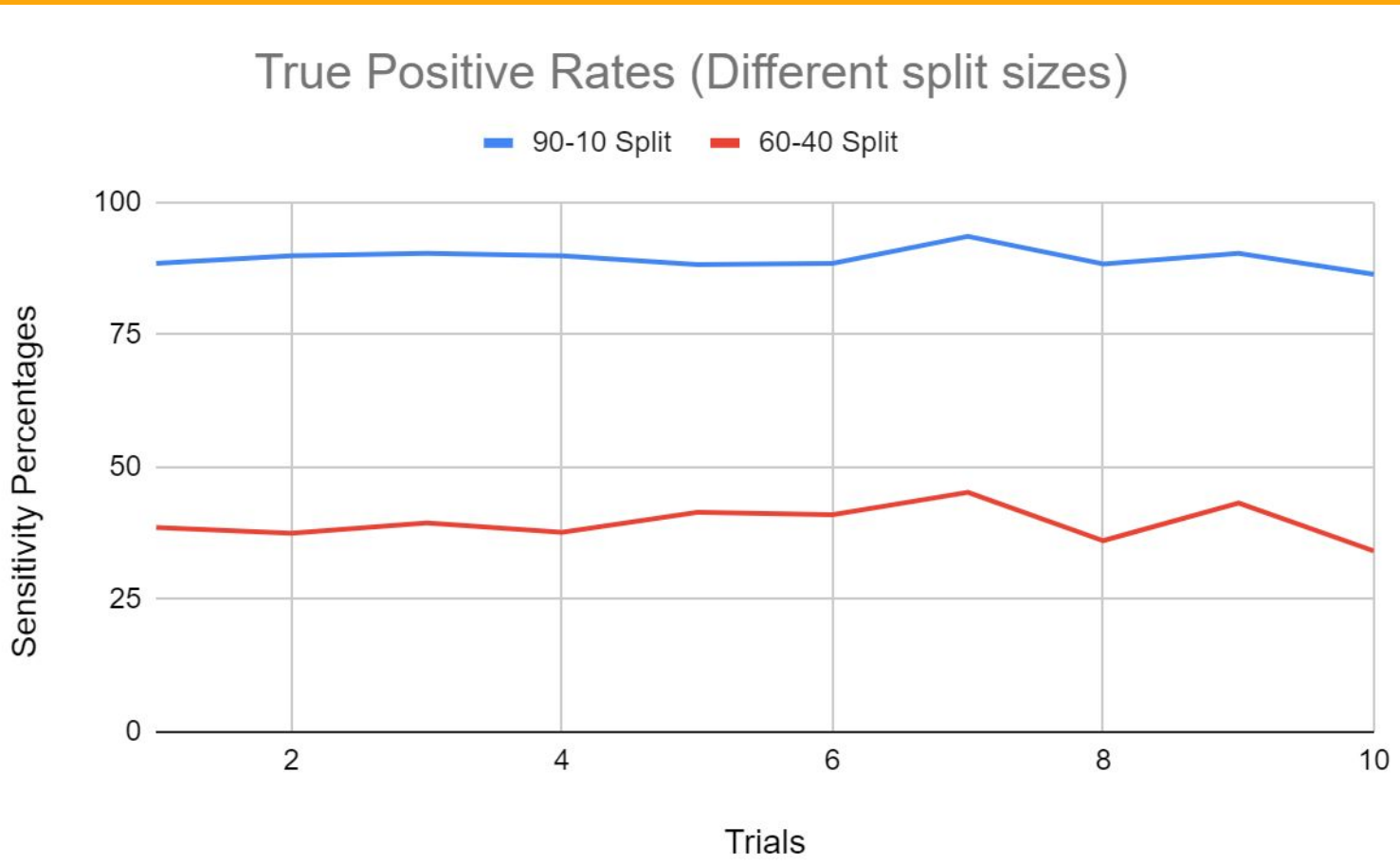


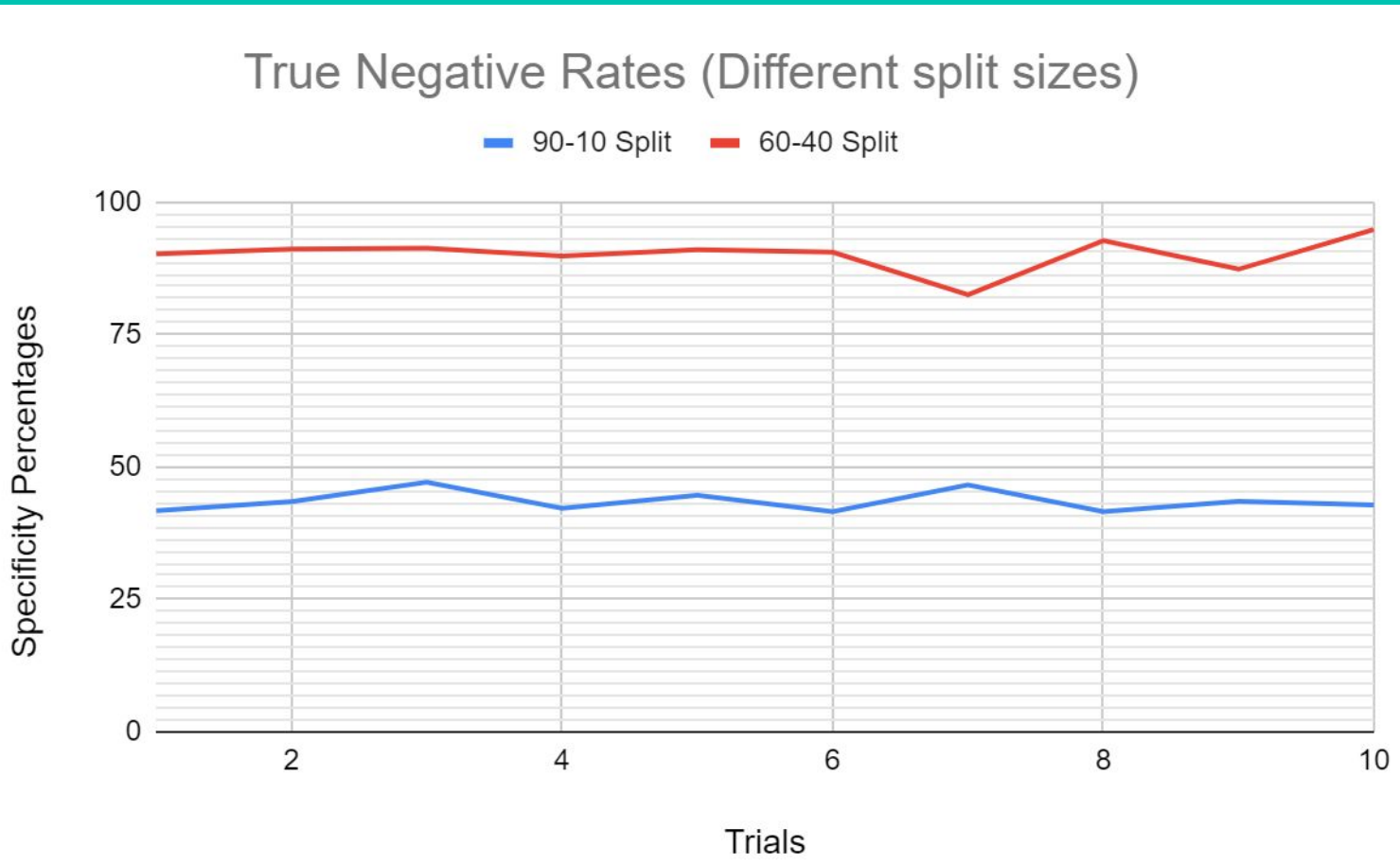
39.4% TP
90.1% TN

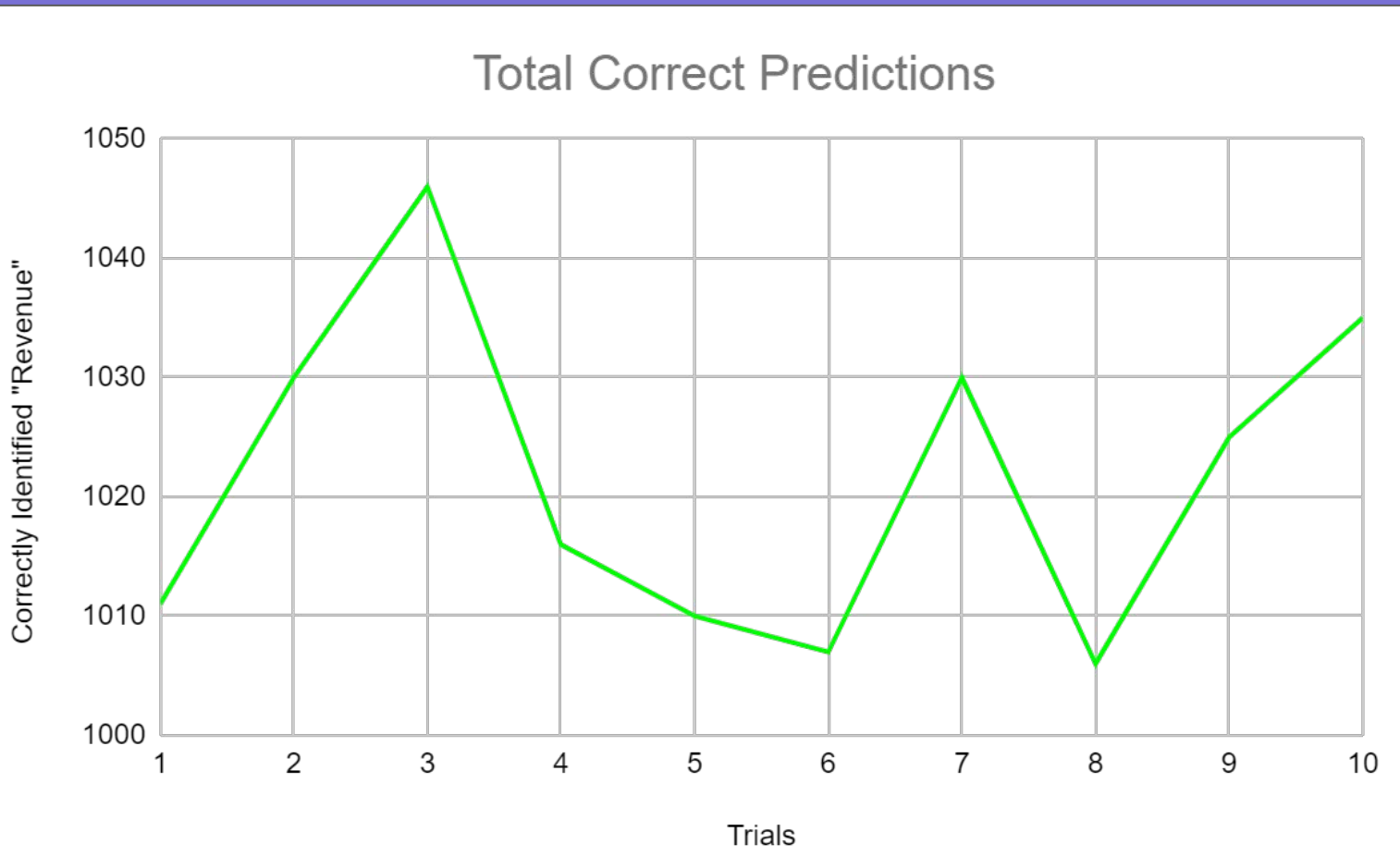
Average accuracy in the 60-40 train-test split dataset

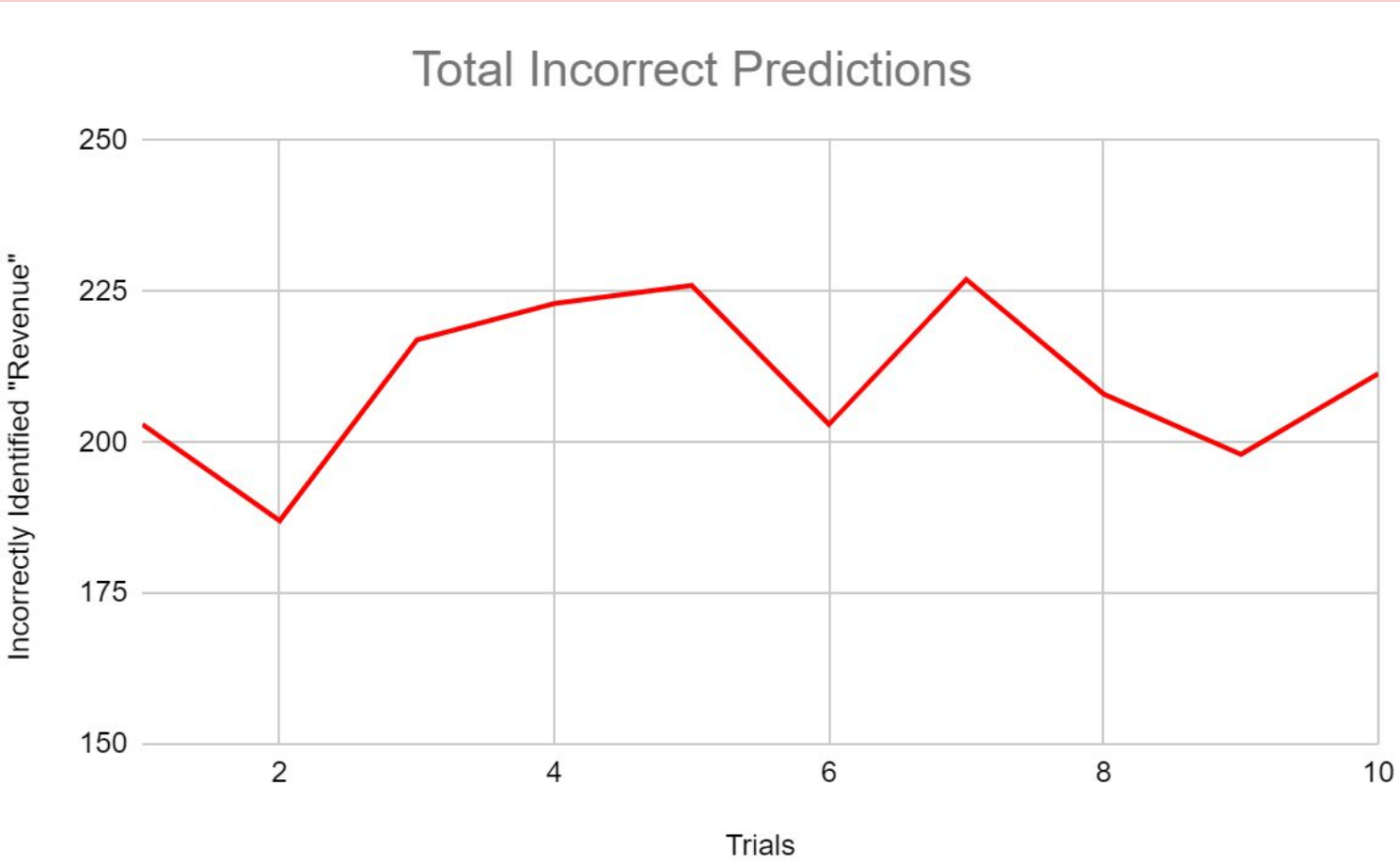
1021 Correct
211 Incorrect

Average correct/incorrect predictions based on K-nearest neighbor algorithm


True Positive Rates (Different split sizes)
90-10 Split
60-40 Split
Sensitivity Percentages
100
75
50
25
0
2
4
6
8
10
Trials


True Negative Rates (Different split sizes)
90-10 Split
60-40 Split
100
75
50
25
0
Specificity Percentages
2
4
6
8
10
Trials

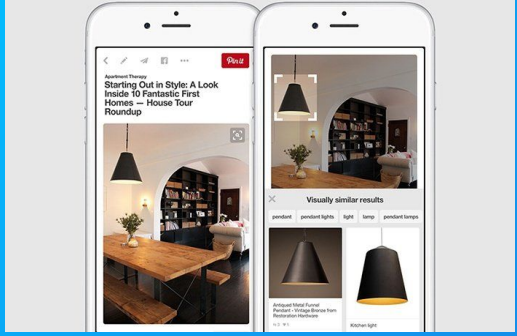

Total Correct Predictions
Correctly Identified "Revenue"
1050
1040
1030
1020
1010
1000
1
2
3
4
5
6
7
8
9
10
Trials


Total Incorrect Predictions
Incorrectly Identified "Revenue"
250
225
200
175
150
2
4
6
8
10
Trials

Overview

Key Takeaways:

- With increasing testing size, the true positive rate decreases indicating a more prominent trend of non buying customers
- Throughout each individual testing size, the true positive and true negative rates stayed fairly consistent.
- Correctly predicted around 4 out of every 5 users' purchasing intents
- At times slow with testing with larger testing sizes but responsive with smaller sizes ("real time")



Future Applications

Multi-device Experience
Personalization across mobile devices, laptops, tablets, etc using AI

Advanced AI Algorithms
Predictive Analytics, Financial Aids using AI

Customer-centric search
Advanced image and video recognition software with customized search options

New Marketing Methods
Content creation, Automated decision-making, Natural language processing

Brand Differentiation
Extensive analysis on websites, social media, to promote one's product over another (generate news, attract traffic, etc)

Ad Campaigns
Optimize target audience, New ad-making strategies