Real-Time Prediction of Online Shoppers' Purchasing Intents Using K-Nearest Neighbor Classification
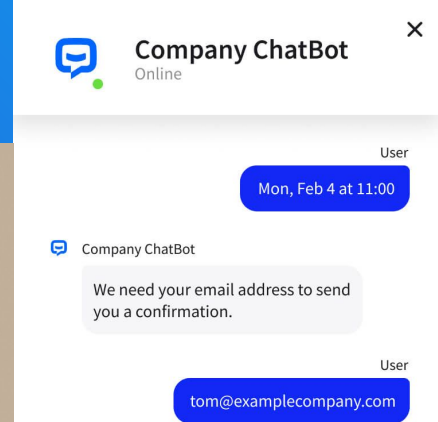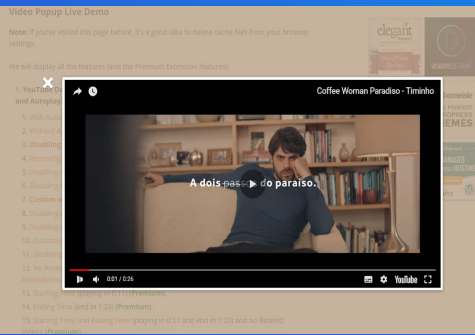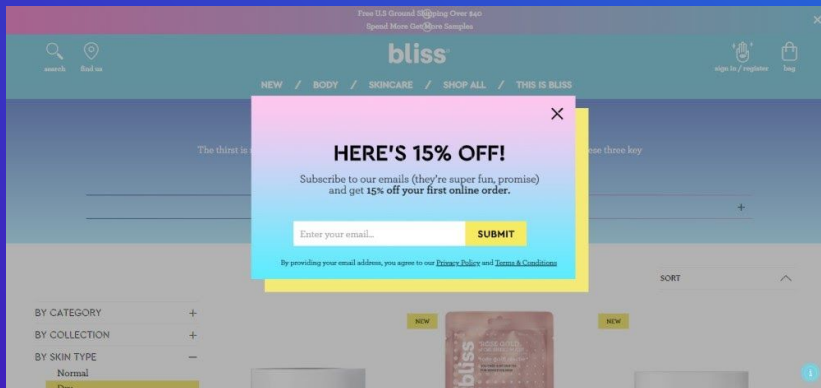
Pranav Tonpe, Plano West Senior High School
AYSI Senior Division

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

Website Link Clicked

Informational
(News articles, research reports, etc.)

Website AI displays promotional media

User is distracted and/or bothered

Product-related
(discounts, pop-up ads, customers)

AI recognizes a purchasing intent from user

Personalized ads capture the attention of user

Administrative
(government websites)

AI deduces a low likelihood of buying a product

May or may not display marketing media

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
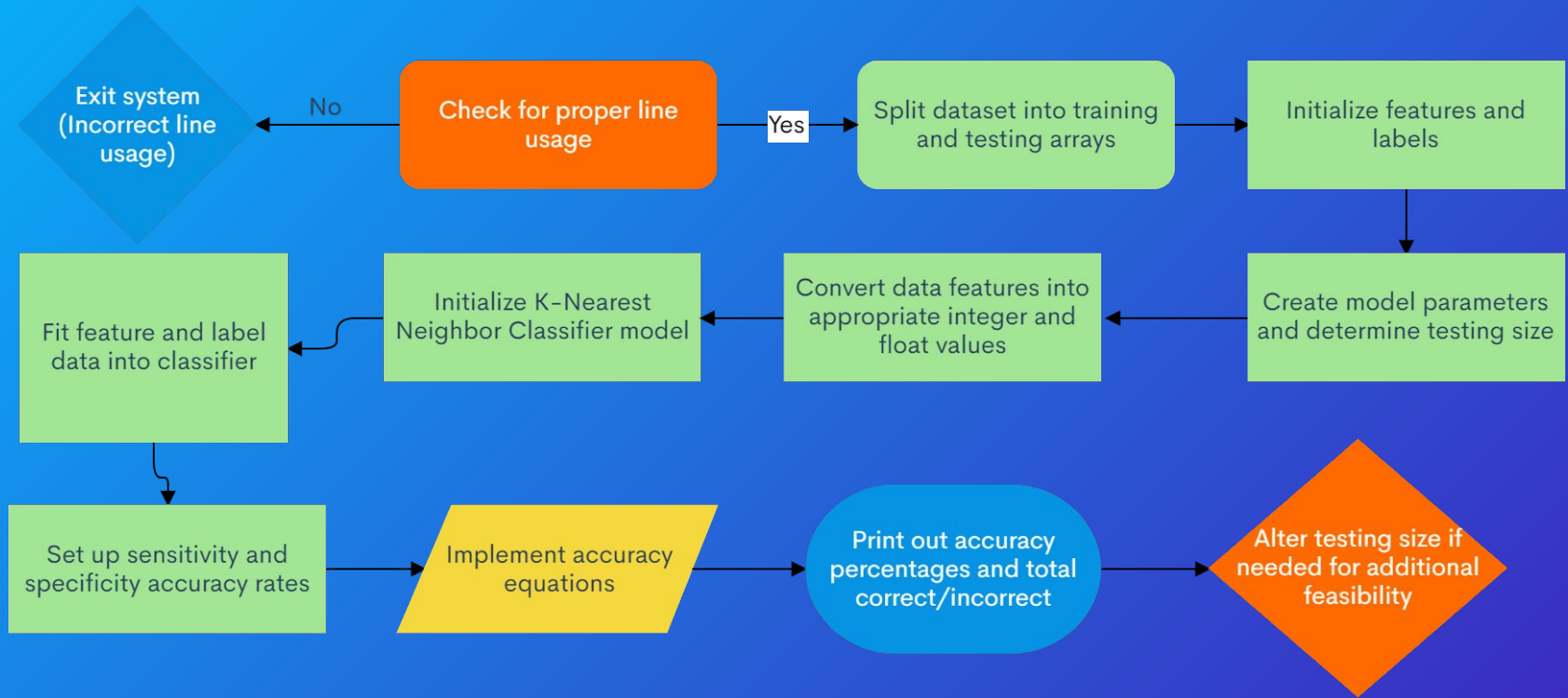
03
Data fitting
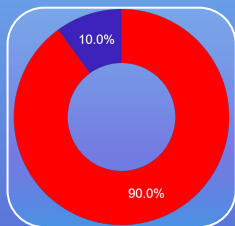Create K-nearest neighbor model and analyze outcomes to make future predictions

04
Accuracy Metrics
Initialize specificity and sensitivity rates and implement their equations

Flowchart:

- **Check for proper line usage** (decision)
  - **No** → **Exit system (Incorrect line usage)**
  - **Yes** → **Split dataset into training and testing arrays**
- **Split dataset into training and testing arrays** → **Initialize features and labels**
- **Initialize features and labels** → **Create model parameters and determine testing size**
- **Create model parameters and determine testing size** → **Convert data features into appropriate integer and float values**
- **Convert data features into appropriate integer and float values** → **Initialize K-Nearest Neighbor Classifier model**
- **Initialize K-Nearest Neighbor Classifier model** → **Fit feature and label data into classifier**
- **Fit feature and label data into classifier** → **Set up sensitivity and specificity accuracy rates**
- **Set up sensitivity and specificity accuracy rates** → **Implement accuracy equations**
- **Implement accuracy equations** → **Print out accuracy percentages and total correct/incorrect**
- **Print out accuracy percentages and total correct/incorrect** → **Alter testing size if needed for additional feasibility**

Ideal Situation

1.

AI classifier recognizes a purchasing intent

2.

Website sends promotional media to lure the user to buy a product

3.

Customer enjoys his/her personalized website with exclusive offers

4.

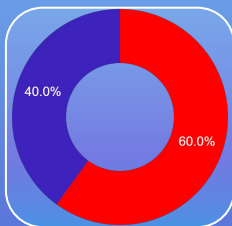A purchase is made → $$$ for company and satisfied customer

Results

89.4% TP
43.5% TN

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

39.4% TP
90.1% TN
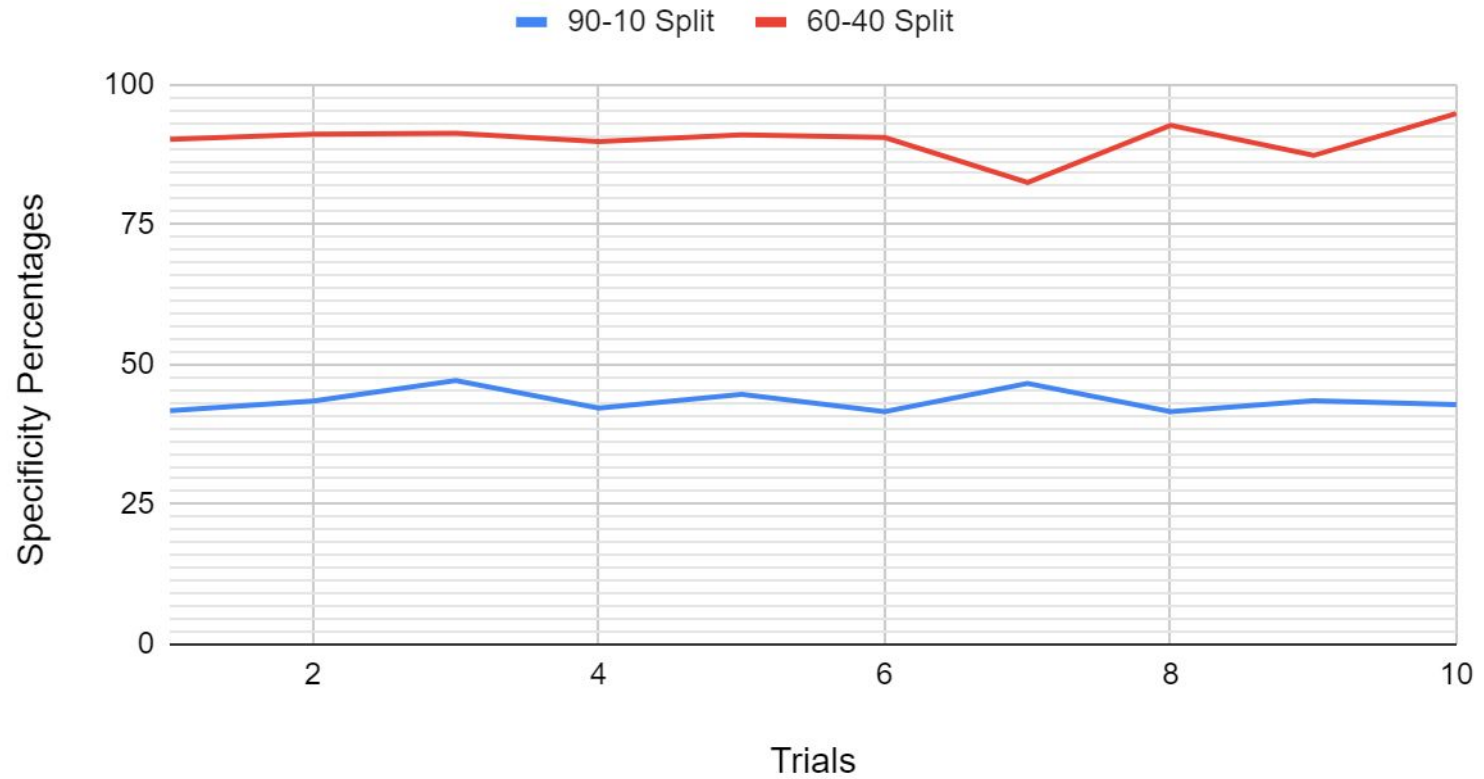
Average accuracy in the
60-40 train-test split
dataset

1021 Correct
211 Incorrect

Average correct/incorrect
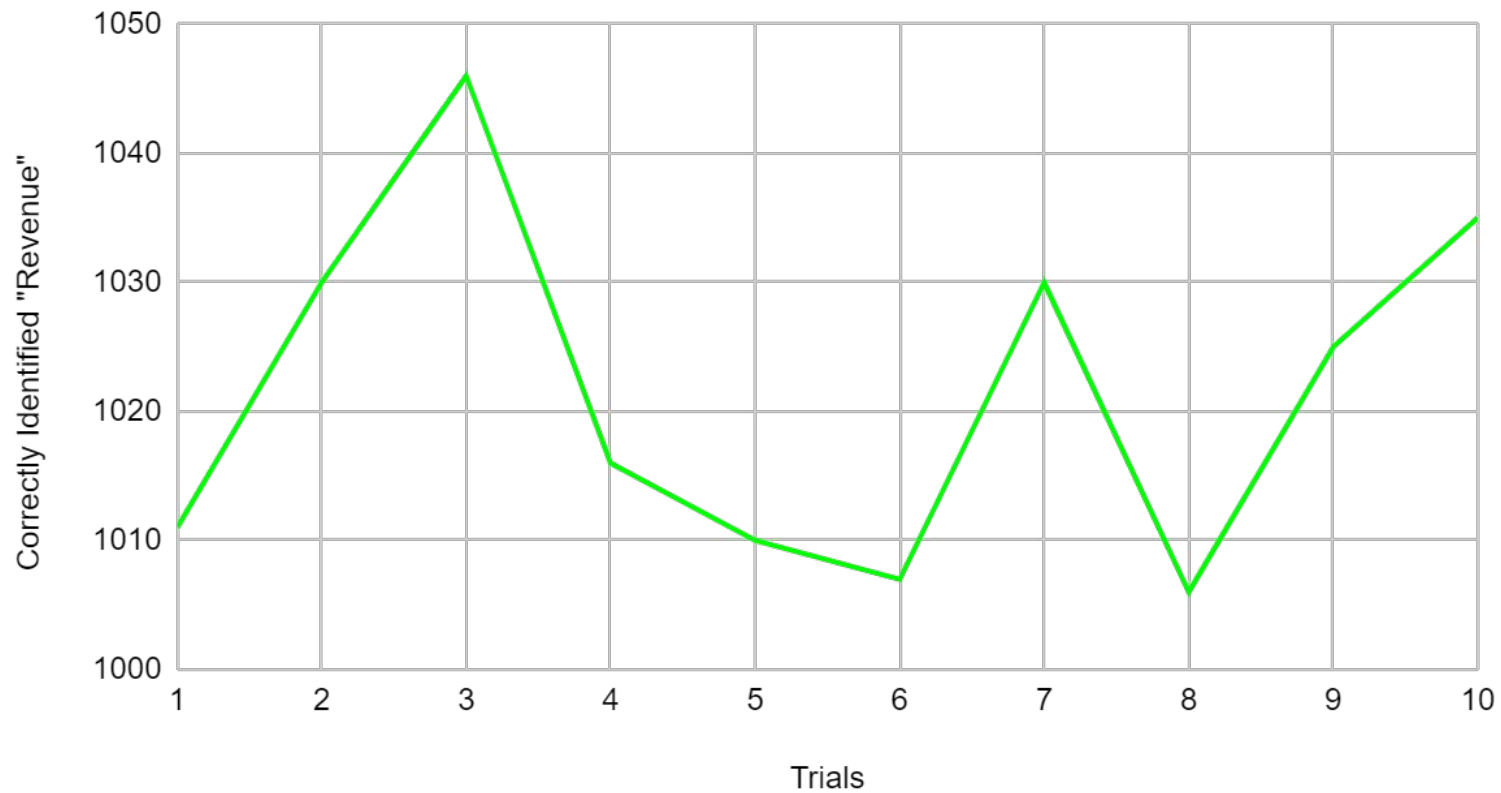predictions based on
K-nearest neighbor algorithm
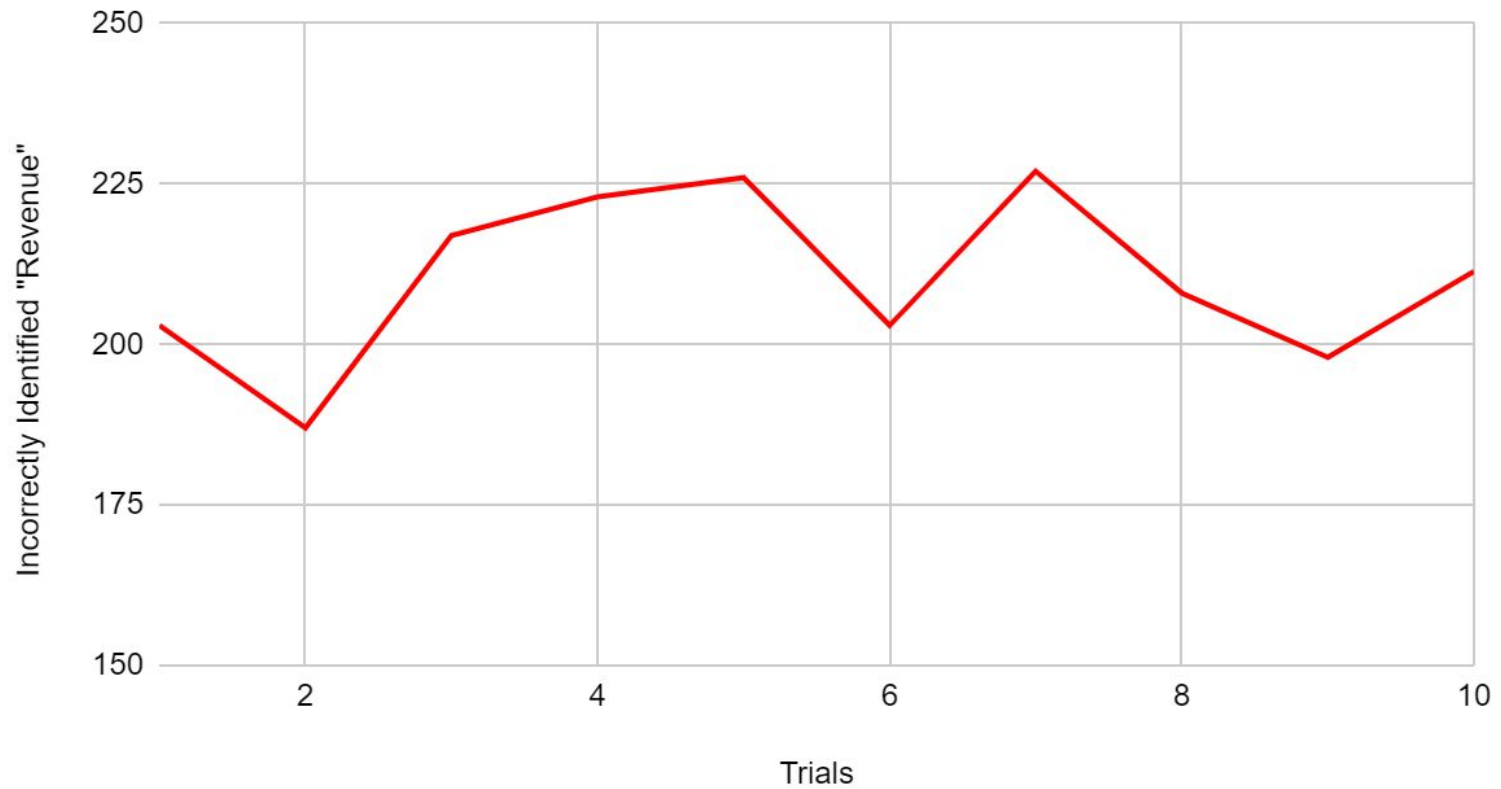
True Positive Rates (Different split sizes)
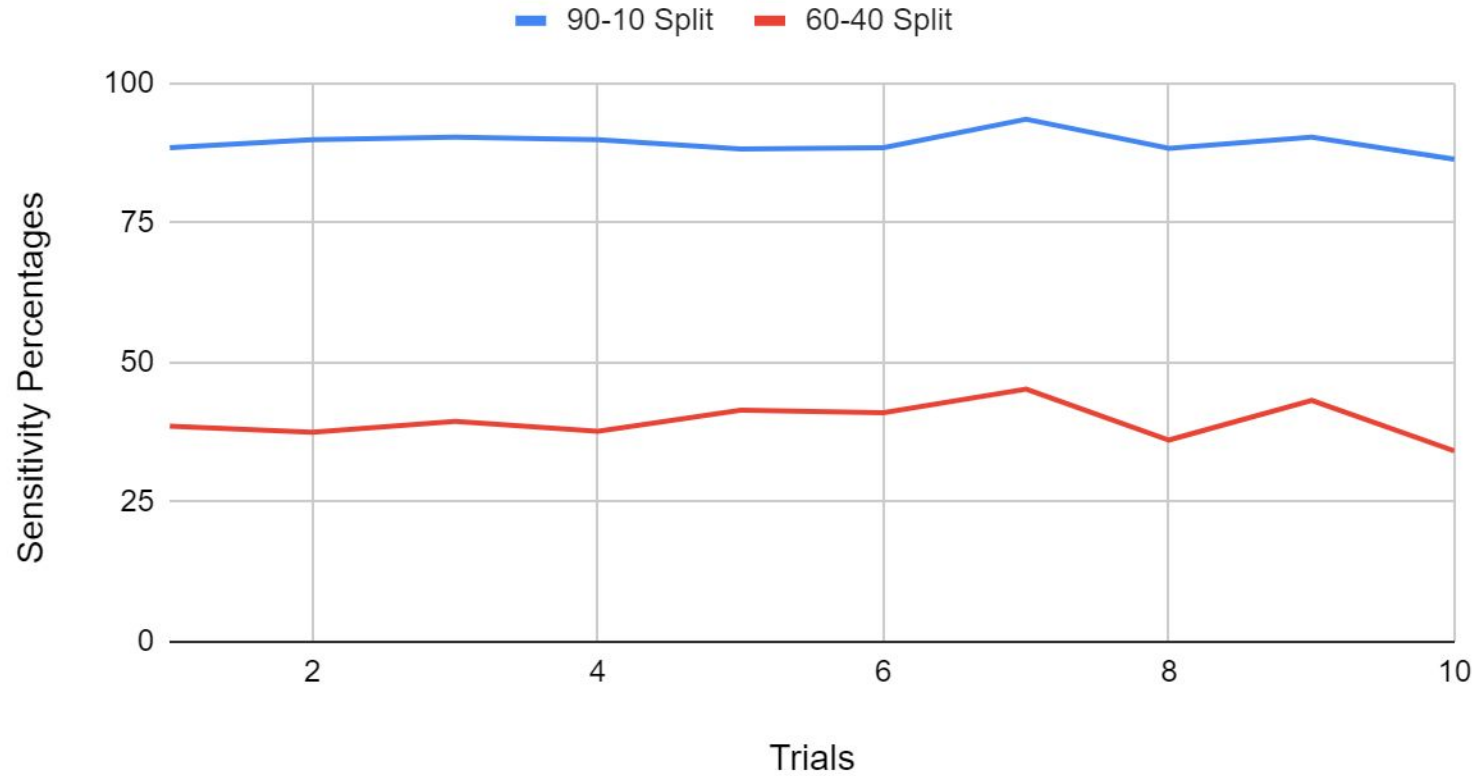
True Negative Rates (Different split sizes)

Total Correct Predictions

Total Incorrect Predictions

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