A smartphone is shown at an angle, displaying a text message on its screen. The message is in a bubble and contains garbled text. The phone is resting on a surface, and the background is a blurred city skyline at sunset or sunrise.

SMS Spam Detection using Machine Learning

This presentation delves into a real-world application of machine learning, focusing on classifying SMS messages as legitimate or spam. We will explore the dataset, tools used, and model performance.



by Vihan Singh



Introduction

The Challenge

Spam messages are a growing nuisance, clogging inboxes and sometimes containing malicious content.

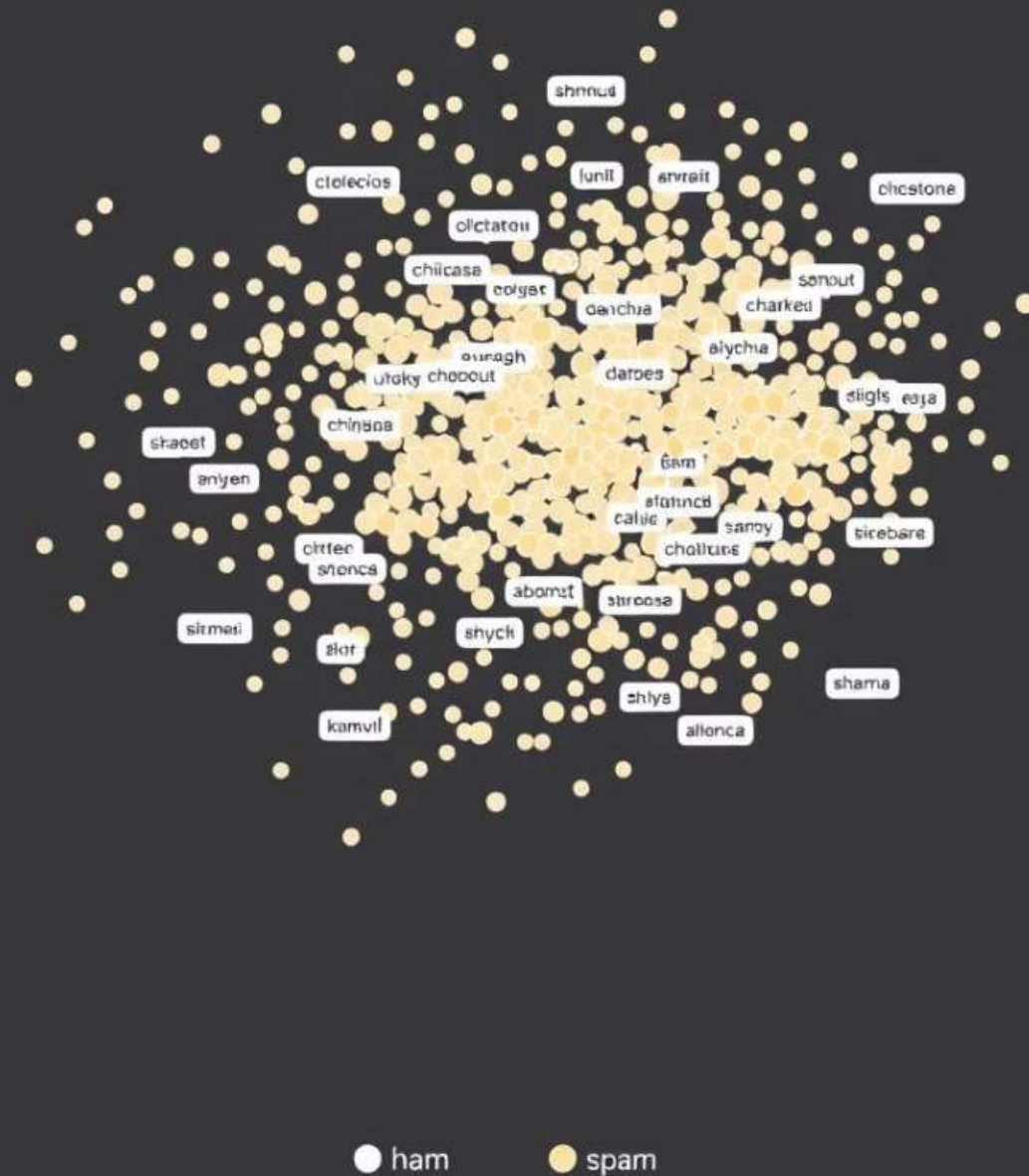
Our Solution

This project aims to develop a machine learning model that effectively identifies and filters spam SMS messages.



SMS e L.SL elltesate rait

setple mot and porifed colhestones



Dataset

Data Source

The SMS Spam Collection
dataset, a widely used
benchmark in text classification.

Data Size

Containing 5,574 samples,
providing sufficient data for
training and testing.

Data Format

Raw SMS text messages, requiring preprocessing for efficient analysis.



Tools and Technologies



Python

Chosen for its rich libraries and ease of use in data science and machine learning.



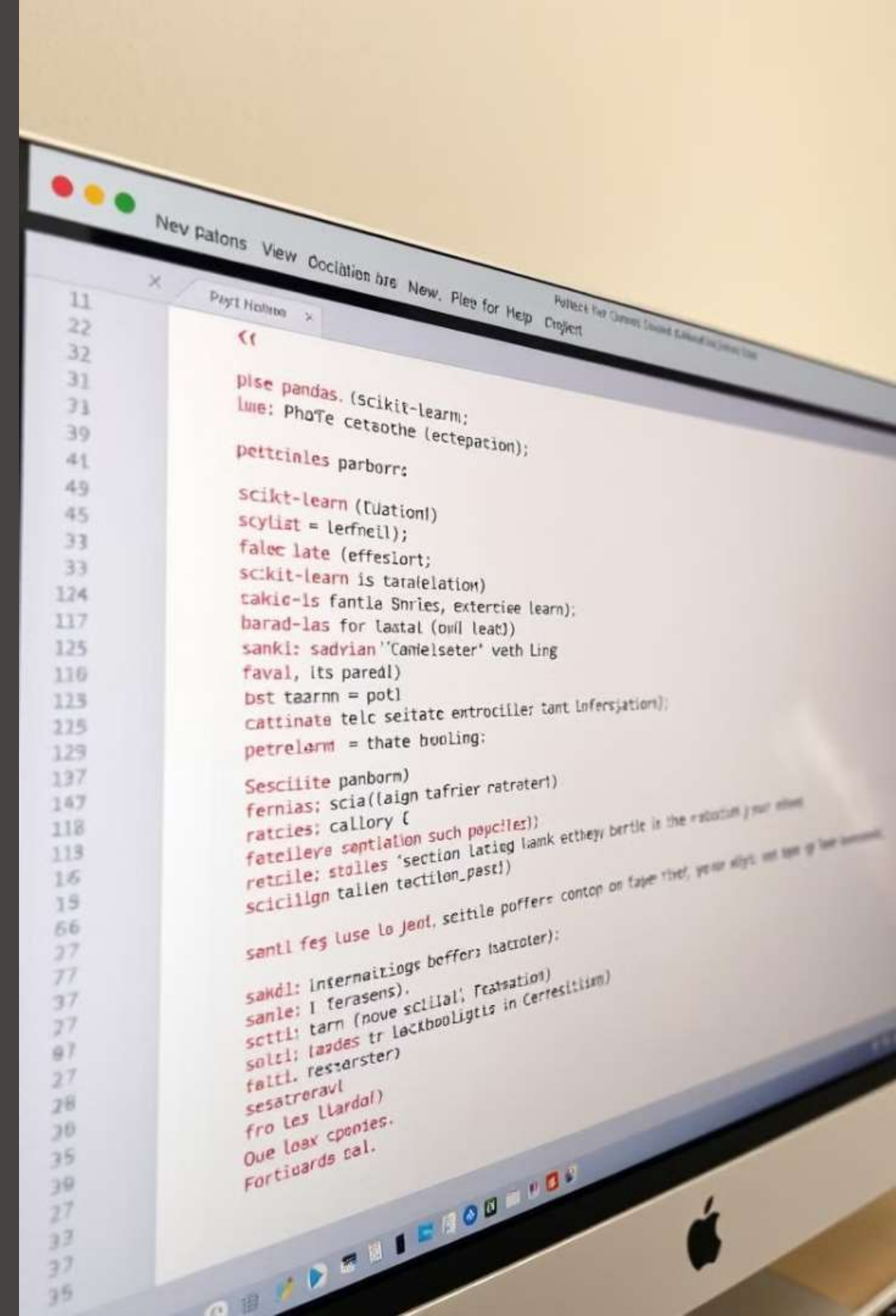
Machine Learning

Algorithms like Naive Bayes, SVM, Logistic Regression, and Random Forest were employed.



Visualization

Libraries like matplotlib and seaborn were used for creating insightful data visualizations.



Snina SMS Deregly



Spa cuninder

D.  ste
ontclar



Catunal
A Micclal

Conpger Steaction

S_i
- traintign -



Deployiag

Workflow

1

Data Preprocessing

Cleaning, tokenization, stop word removal, stemming to prepare the text data.

2

Feature Engineering

Converting text into numerical representations using Bag-of-Words and TF-IDF.

3

Train-Test Split

Dividing the dataset into 80% training data and 20% testing data for model evaluation.



Model Development

1

1. Naive Bayes

A probabilistic approach based on Bayes' theorem to classify messages as ham or spam.

2

2. Support Vector Machine (SVM)

A powerful algorithm that finds a hyperplane to separate the data points into different classes.

3

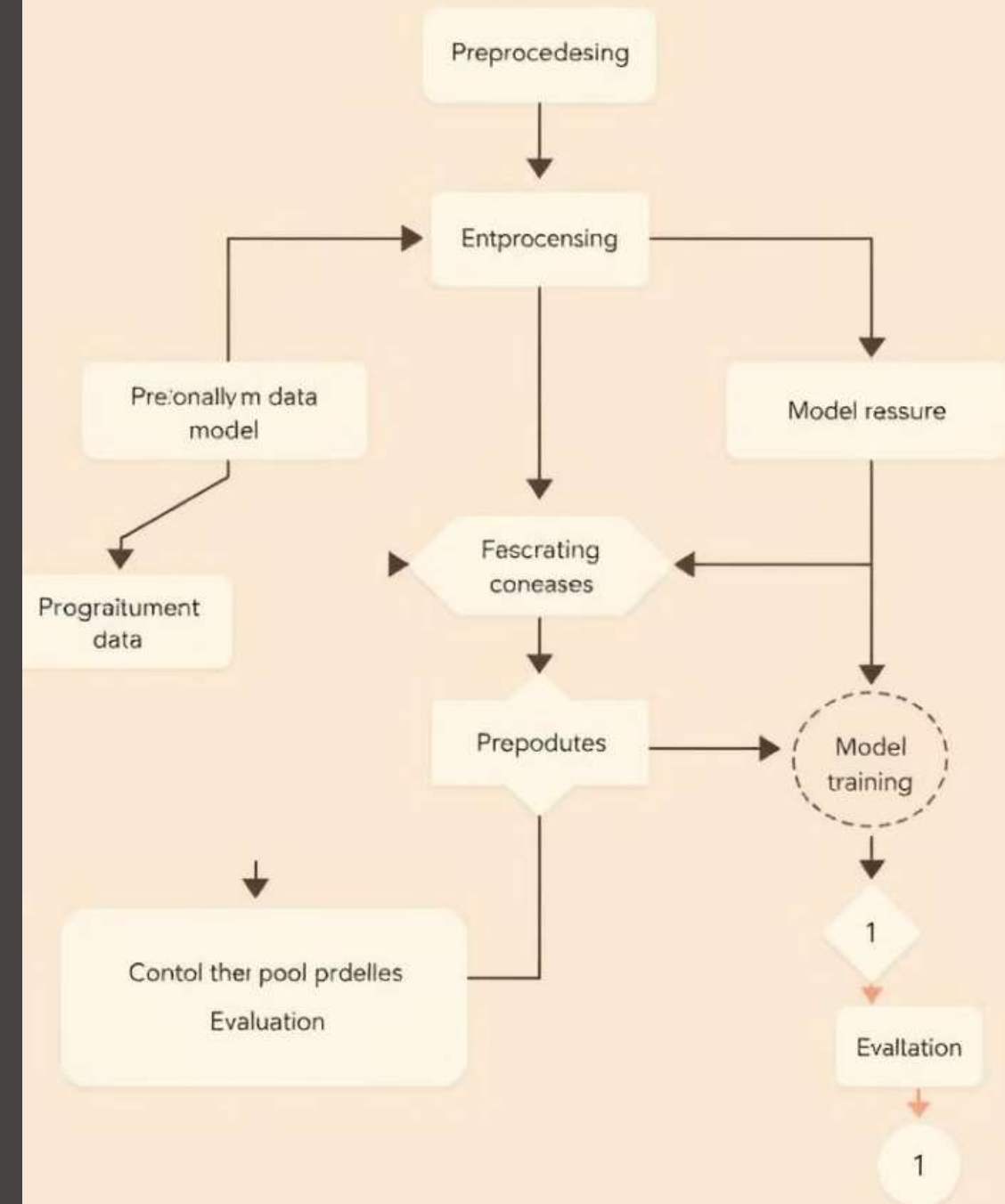
3. Logistic Regression

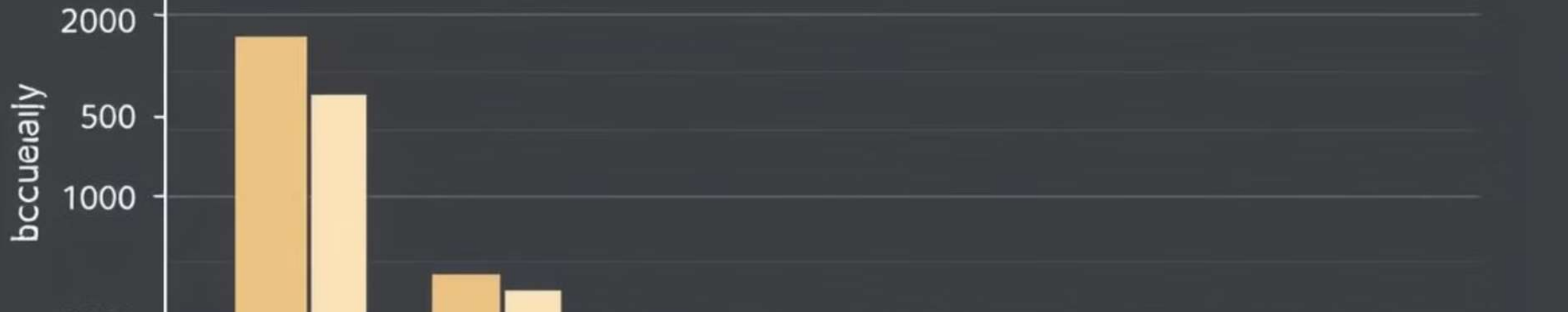
A statistical method that predicts the probability of a message belonging to a particular class.

4

4. Random Forest

An ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting.





Results

97%

Accuracy

The best-performing model achieved a high accuracy rate, indicating its effectiveness in identifying spam.

95%

Precision

The model accurately identified spam messages, minimizing false positives.

96%

Recall

The model successfully identified most spam messages, minimizing false negatives.



Conclusion

1

High Performance

The spam detection model effectively classifies SMS messages with high accuracy, precision, and recall.

2

Effective Preprocessing

Data preprocessing and feature engineering play a crucial role in improving model performance.

3

Practical Application

The model can be implemented in real-world applications to protect users from spam messages.





Thank You

We encourage questions and feedback on this project. Contact me for further discussion or collaboration.

