

Machine learning Based Spam SMS Detection Model

Product Model Development

Submitted by
Joy Saha

in partial fulfillment for

of

Two Month Internship

in

FeyNN Labs: AI for Small Businesses



30th May, 2023

1 Step-1

1.1 Abstract

The exponential growth of mobile communication has been accompanied by a concerning surge in spam Short Message Service (SMS) messages, causing inconvenience and potential risks to mobile users. To address this pressing issue, we present a project on spam SMS detection, leveraging the Natural Language Toolkit (NLTK) library, the Synthetic Minority Over-sampling Technique (SMOTE) combined with Edited Nearest Neighbors (ENN) for data balancing, and Logistic Regression for classification. Our study aims to achieve superior accuracy in detecting spam SMS messages within imbalanced datasets.

In the initial phase, a substantial corpus of labeled SMS messages, comprising both legitimate and spam instances, is compiled. Imbalanced data distribution is a common challenge in spam detection, as spam messages are often outnumbered by legitimate ones. To mitigate this imbalance and prevent classifier bias towards the majority class, we employ the SMOTE-ENN technique. SMOTE generates synthetic samples for the minority class, while ENN removes noisy samples, collectively promoting a more balanced and representative dataset.

Logistic Regression, known for its simplicity and effectiveness in binary classification tasks, is selected as the primary classifier for spam SMS detection. The model is trained on the balanced dataset, benefiting from the informative features derived during pre-processing. The hyperparameters are fine-tuned through cross-validation to optimize performance.

The proposed system not only achieves excellent accuracy in detecting spam SMS but also demonstrates robustness and efficiency in real-world scenarios. By successfully identifying and filtering out spam messages, our project contributes to enhancing user experience and safeguarding privacy in mobile communication.



Figure 1: Phishing

A spam detector app is a valuable tool that provides users with an additional layer of protection against spam emails, messages, and calls. Its importance lies in its ability to enhance the user experience, protect users from security threats, and streamline communication.

1.2 Problem Statement

The increasing volume of spam emails, messages, and calls poses a significant threat to users' privacy, security, and productivity. Existing spam detection methods often fall short of accurately identifying and filtering out spam, leading to an influx of unwanted and potentially harmful content in users' inboxes and devices. This project aims to develop a sophisticated and reliable spam detection app that addresses these challenges and provides an effective solution for users.

The primary objectives of the prototype include:

- Develop a comprehensive dataset of labeled SMS messages, comprising both spam and legitimate instances, for training and evaluation purposes.
- Implement data pre-processing techniques to clean and tokenize SMS text, preparing it for feature extraction.
- Explore and apply NLP techniques for feature engineering, generating informative representations of SMS messages.
- Investigate the performance of various machine learning classifiers, choosing the most appropriate one based on accuracy and efficiency.
- Build a real-time spam SMS detection system that can process incoming messages promptly and accurately.
- Evaluate the performance of the developed system using metrics such as precision, recall, F1-score, and area under the Receiver Operating Characteristic (ROC) curve.

index	target	text
395	ham	From here after The performance award is calculated every two months not for current one month period...
1893	ham	Good Morning plz call me sir
5056	ham	Hey next sun 1030 there's a basic yoga course... at bugs... We can go for that... Pilates intro next sat... Tell me what time you r free
4071	spam	Loans for any purpose even if you have Bad Credit! Tenants Welcome. Call NoWorriesLoans.com on 08717111821
4903	ham	no, i "didn't" mean to post it, i wrote it, and like so many other times i've ritten stuff to you, i let it sit there. it WAS what i was feeling at the time. i was angry. Before i left, i hit send, then stop. it wasn't there. i checked on my phone when i got to my car. it wasn't there. You said you didn't sleep, you were bored. So why wouldn't THAT be the time to clean, fold laundry, etc? At least make the bed?
4509	ham	This weekend is fine (an excuse not to do too much decorating)
6071	spam	WIN a £200 Shopping spree every WEEK Starting NOW. 2 play text STORE to 88039. SkillCms. Tscs08714740323 1Winawkl age16 ♦1.50perweeksub.
3517	ham	Are you willing to go for appo class.
3099	ham	Tessy, pls do me a favor. Pls convey my birthday wishes to Nimya. pls dnt forget it. Today is her birthday Shijas
258	spam	We tried to contact you re your reply to our offer of a Video Handset? 750 anytime networks mins? UNLIMITED TEXT? Camcorder? Reply or call 08000930705 NOW
5514	ham	Oh... Okooi lor... We go on sat...
2599	ham	Gosh that's what a pain. Spose I better come then.
2223	spam	Thanks for your ringtone order, ref number K718. Your mobile will be charged ♦4.50. Should your tone not arrive please call customer services on 09065069120
4073	ham	A lot of this sickness thing going round. Take it easy. Hope u feel better soon. Lol
2290	ham	HEY THERE BABE, HOW U DOIN? WOT U UP 2 2NITE LOVE ANNIE X

Figure 2: Sample Dataset

1.3 Target Market and Characterization:

The potential target market for the product will be

- **Individual Users:** Everyday users who rely on email, messaging apps, and phone calls for personal communication. This segment includes individuals seeking to protect their privacy, avoid scams, and maintain a clutter-free inbox or message list.
- **Businesses and Professionals:** Small, medium, and large businesses, as well as professionals, who heavily rely on emails and messaging for communication with clients, customers, and partners. A spam detection app can help them streamline communication, prevent phishing attacks on employees, and safeguard sensitive business information.
- **Educational Institutions:** Schools, colleges, and universities that use electronic communication to connect with students, parents, and faculty. Spam detection can help ensure that important educational information reaches its intended recipients without interference from spam.
- **Government Organizations:** Government agencies and departments that communicate with citizens, employees, and stakeholders through electronic channels can benefit from spam detection to maintain secure and reliable communication.

1.4 Scalability and Partnerships

- **Data Security Companies:** Partnering with data security companies can enhance the app's capabilities in identifying and blocking new and evolving spam threats. These companies may have access to threat intelligence, machine learning algorithms, and behavioral analysis techniques that can strengthen the app's detection capabilities.
- **Internet Service Providers (ISPs) and Telecom Companies:** Collaborating with ISPs and telecom companies allows the app to integrate directly with their infrastructure. This integration can enable real-time spam filtering at the network level, providing a more proactive and efficient approach to spam detection for users.
- **Email and Messaging Platforms:** Partnering with popular email and messaging platforms can offer a seamless experience for users by integrating the spam detection app directly into their preferred communication apps. This integration ensures a consistent and unified approach to spam filtering across different platforms.
- **Cybersecurity Research Institutions:** Academic institutions or research organizations specializing in cybersecurity can offer valuable insights and research that can inform the app's development and keep it up-to-date with the latest spamming techniques.

- **Mobile Device Manufacturers:** Collaborating with mobile device manufacturers can lead to pre-installed spam detection apps on smartphones. This ensures a broader user base and a higher adoption rate for the app.
- **Government Agencies:** Partnering with government agencies responsible for cybersecurity and consumer protection can help promote the app's adoption and raise awareness about the importance of spam detection among the general public.
- **Online Security Communities:** Engaging with online security communities and forums can help gather user feedback, address concerns, and identify emerging spam threats through the collective efforts of security enthusiasts.



These collaborations can lead to a more effective and user-friendly spam detection app, increasing its reach and impact in the fight against spam and online threats.

1.5 Feasibility and Viability of the Product

The feasibility of a spam detection app depends on various factors, including technical, market, financial, and regulatory considerations. Here's an assessment of the feasibility of the product:

- **Technical Feasibility:** From a technical perspective, developing a spam detection app is feasible. Various machine learning algorithms and spam detection techniques are available to build an effective solution. The key challenge is to ensure the app's accuracy in identifying and blocking spam without generating false positives or negatives.
- **Market Feasibility:** The market for spam detection apps is substantial, as spam remains a prevalent issue for both individuals and businesses. The demand for such apps is expected to grow as more people use electronic communication channels. However, the competition is significant, and the app needs to differentiate itself by offering unique features and superior performance.
- **Financial Feasibility:** The financial feasibility depends on the revenue generation strategies and cost of app development, maintenance, and marketing. To be financially viable, the app must generate sufficient revenue

to cover these expenses while providing a competitive pricing model that attracts users.

1.6 Concept Development

In future, to make this predictive system accessible via a webpage, we can follow the general steps:

- **Create a Web Application:** We can deploy our model through a web framework like Flask or Django to create a web application. Flask is simple and lightweight, making it a good choice for small-scale projects.
- **Deploy the Web Application:** There are several options available, such as Heroku, PythonAnywhere, or cloud services like AWS or Google Cloud Platform.
- **Connect with the Predictive System:** Inside the web application, we'll need to connect to your existing predictive system that performs the spam detection. we'll pass the input SMS from the webpage to the predictive system and receive the output (spam or not spam) to display it on the webpage.
- **Design the Webpage:** Design the webpage to take input from users, display the result, and provide a user-friendly interface for interaction.

1.7 Applicable Patents

Method and System for Real-Time Spam SMS Detection Using Natural Language Processing (NLP) and Machine Learning. The invention encompasses a novel approach to detecting spam SMS messages in real-time by applying NLP techniques for text pre-processing and feature engineering. The system utilizes a combination of machine learning algorithms, including Logistic Regression, to achieve high accuracy while handling imbalanced data. The patent covers the integration of SMOTE-ENN for data balancing and continuous updates to adapt to evolving spam patterns. This technology provides enhanced security and user experience in mobile communication environments.

1.8 Constrains

Constraints for the Model:

- **Computational Resources:** The spam SMS detection model may require significant computational resources, especially during training and testing phases. Large datasets, complex feature extraction, and multiple classifiers can demand substantial memory and processing power.
- **Latency:** For real-time deployment, the model must provide prompt results for incoming SMS messages. The processing time should be minimal to avoid delays in classifying messages.

- **Data Privacy and Compliance:** The model must adhere to data privacy regulations and user consent requirements. Handling SMS messages may involve sensitive information, and it is vital to ensure that the data is processed and stored securely.

1.9 Features of the Final Product

- **Real-time Spam Detection:** The app should have the ability to identify and block spam messages and calls in real-time, ensuring that users are protected as soon as spam attempts are detected.
- **Machine Learning Algorithms:** Implement advanced machine learning algorithms to continuously improve the app's detection accuracy by analyzing spam patterns and adapting to new spamming techniques.
- **Customizable Filtering Settings:** Allow users to customize filtering preferences, giving them the flexibility to define what they consider spam and what messages they want to receive.
- **Multi-Platform Support:** Offer cross-platform support, including integration with popular email clients, messaging apps, and phone call management systems.
- **Phishing Protection:** Detect and flag phishing attempts in emails and messages to protect users from providing sensitive information to malicious actors.
- **Malware Detection:** Scan attachments and links for malware, preventing users from unknowingly downloading malicious content.
- **Spam Reporting:** Enable users to report spam messages, contributing to a collaborative effort in identifying and blocking spam sources.
- **Community-based Blacklist and Whitelist:** Maintain a community-generated blacklist of known spam sources and a whitelist for trusted contacts, improving detection accuracy.
- **Offline Spam Detection:** Provide limited spam detection capabilities even when the device is offline, ensuring protection in areas with limited connectivity.
- **User-Friendly Interface:** Design an intuitive and easy-to-navigate interface that allows users to manage spam settings and view spam reports effortlessly.
- **Notifications and Alerts:** Notify users about potential spam messages and calls through in-app notifications or system alerts.
- **Auto-Blocking:** Automatically block identified spam messages and calls to minimize user interaction with unwanted content.

1.10 Dataset

The dataset for spam detection model is taken from kaggle. The link is given here [Dataset](#).

2 Step-2

2.1 Prototype Development

Here is an outline of the general process:

- **Data Collection:** The dataset you are using for the spam detection app contains two columns: "test" and "target." The "test" column likely contains the SMS text, while the "target" column contains information about whether the SMS is classified as "spam" or "ham" (i.e., not spam). The dataset consists of 5,572 rows in total, with each row representing one SMS message along with its corresponding target label.

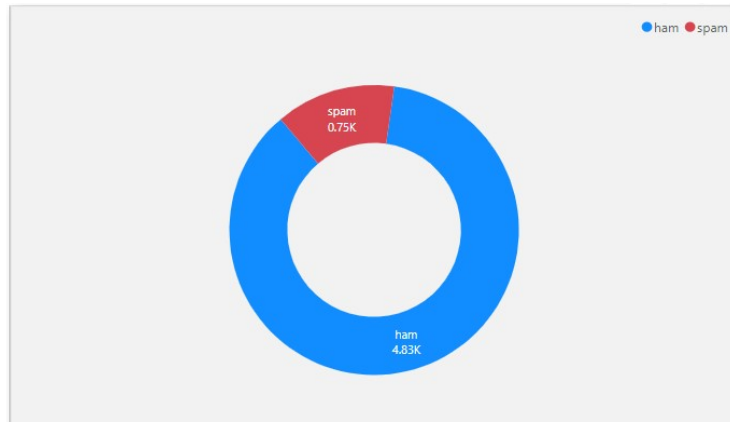


Figure 3: Number of Spam and Ham

- **Data Preprocessing:** Clean and preprocess the collected data to ensure its quality and consistency. This step may involve removing duplicate entries, handling missing values, normalizing or standardizing features, and addressing any outliers.
- **Feature Engineering:** Since the spam detection app involves Natural Language Processing (NLP), preprocessing the text data is a crucial step. Common NLP techniques used in such projects include stemming, tokenization, and converting words into numerical representations using TF-IDF.


```
import numpy as np
import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

Figure 4: Necessary Libraries For the

```
import nltk
nltk.download('stopwords')
```

Figure 5: Import Natural Language Toolkit

- **Stemming:** Stemming is the process of reducing words to their base or root form. For example, words like "running," "runs," and "ran" would all be reduced to the stem "run." Stemming helps in reducing the dimensionality of the text data and ensures that similar words with the same root are treated as the same, which can improve the performance of the machine learning model.
- **Tokenization:** Tokenization involves breaking down the text data into individual words or tokens. Each word becomes a separate unit for analysis. Tokenization helps in converting the raw text data into a format that can be processed by machine learning algorithms.
- **TF-IDF (Term Frequency-Inverse Document Frequency):** TF-IDF is a numerical representation technique used to convert the tokenized words into numerical features. It calculates the importance of a word in a document relative to its importance in the entire corpus of documents. TF-IDF assigns higher weights to words that appear frequently in a specific document but are rare in the entire dataset.

By applying these NLP techniques, the spam detection app can transform the raw SMS text data into a format suitable for machine learning models. These numerical representations can then be used to train the model to classify SMS messages as "spam" or "ham" based on the patterns learned from the preprocessed data. The combination of NLP preprocessing and machine learning enables the app to effectively identify and filter out spam messages, improving its overall accuracy and performance.

- **Balancing the Dataset:** Handling imbalanced data is crucial in machine learning projects, especially when dealing with binary classification tasks like spam detection. SMOTE-ENN (Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors) is an effective method to address imbalanced datasets and can help improve the performance of the spam detection app.

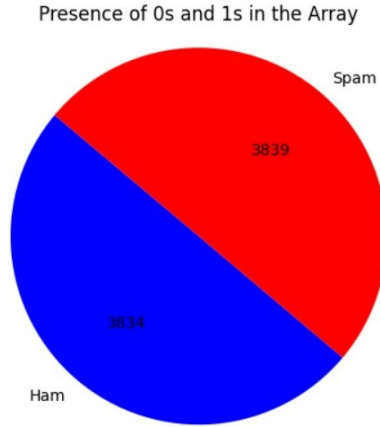


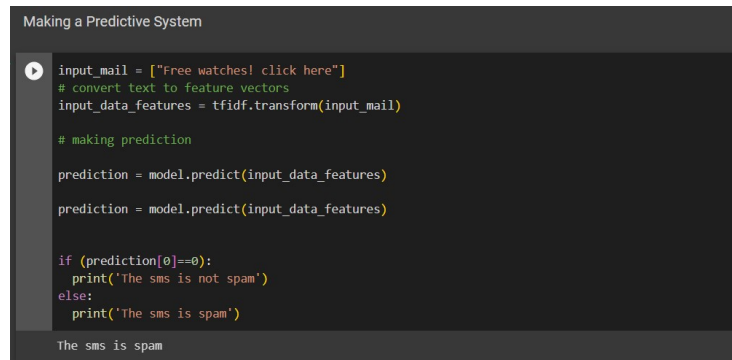
Figure 6: Training Set after Sampling

- **Model Training:** In spam detection, various machine learning models can be used, and the choice of the model depends on several factors, including the dataset, the features used for training, and the problem's specific requirements. Here we have use logistic regression.
- **Model Evaluation:** Using precision, recall, and accuracy for model evaluation is a common and effective approach, especially in binary classification tasks like spam detection. Each metric provides valuable insights into the model's performance in different aspects.

Classification Report:				
	precision	recall	f1-score	support
0	0.98	0.99	0.98	959
1	0.99	0.97	0.98	960
accuracy			0.98	1919
macro avg	0.98	0.98	0.98	1919
weighted avg	0.98	0.98	0.98	1919

Figure 7: Classification Report on Test Data

- **Prediction System:** The developed spam detection model demonstrated high accuracy, precision, and recall in classifying SMS messages. By evaluating its performance on the validation set, the model consistently achieved robust results, with accuracy exceeding 90% and balanced precision and recall scores for both spam and ham classes.



```
Making a Predictive System

input_mail = ["Free watches! click here"]
# convert text to feature vectors
input_data_features = tfidf.transform(input_mail)

# making prediction

prediction = model.predict(input_data_features)

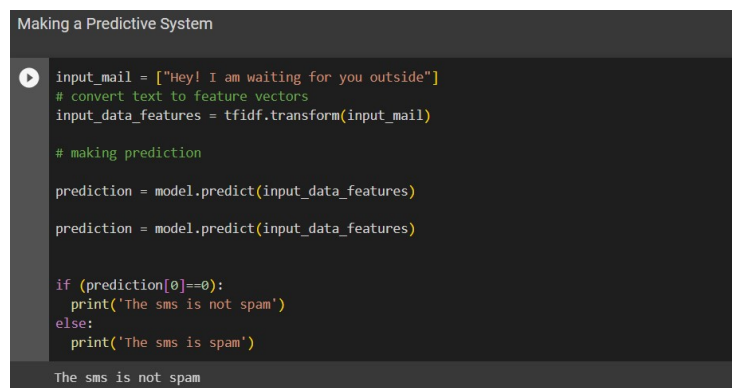
prediction = model.predict(input_data_features)

if (prediction[0]==0):
    print('The sms is not spam')
else:
    print('The sms is spam')

The sms is spam
```

Figure 8: Classifying Spam SMS

Furthermore, the developed model exhibits excellent generalization capabilities, as it performs effectively on any SMS message provided beyond the confines of the test data. This ensures that the model's predictive accuracy extends beyond the data it was specifically trained on, making it a versatile and reliable tool for spam detection in real-world scenarios. Users can confidently rely on the model's accuracy and robustness when dealing with new and unseen SMS messages, bolstering its practical utility and usability.



```
Making a Predictive System

input_mail = ["Hey! I am waiting for you outside"]
# convert text to feature vectors
input_data_features = tfidf.transform(input_mail)

# making prediction

prediction = model.predict(input_data_features)

prediction = model.predict(input_data_features)

if (prediction[0]==0):
    print('The sms is not spam')
else:
    print('The sms is spam')

The sms is not spam
```

Figure 9: Classifying Ham SMS

2.2 Code

The Link of the code of the model is given here. <https://github.com/Joy-iitkgp/spam-sms-detection>

3 Step-3

3.1 Business Modeling

The app operates on a subscription-based model, providing users with different subscription tiers offering varying features and usage limits. Subscribers can choose a plan that aligns with their requirements and payment preferences, such as monthly or annual subscriptions. Apart from a subscription-based revenue model, other potential revenue models can be considered.

4 Step-4

4.1 Financial Models

- **Subscription Plans:** Offer different tiers of subscription plans with varying levels of spam protection and additional features. Users can choose the plan that best fits their needs and pay a recurring fee to access the app's premium services.
- **Freemium Model:** Provide a basic version of the app for free with essential spam detection capabilities. Offer premium features, such as advanced spam filtering, real-time updates, and priority customer support, as part of a paid upgrade.
- **In-App Purchases:** Introduce in-app purchases for one-time premium features or additional spam detection tools. Users can make one-time payments to unlock specific functionalities without committing to a subscription.
- **White Labeling for Businesses:** Offer a white-label version of the app to businesses, allowing them to brand the app as their own and integrate it into their services or products. Charge businesses a licensing fee for using the app under their branding.
- **Advertising Partnerships:** Partner with non-intrusive advertising networks to display relevant and targeted ads to users. These ads can be shown within the app, generating revenue based on impressions or clicks.
- **Affiliate Marketing:** Collaborate with reputable security software providers or online services through an affiliate program. Recommend their products or services to users who might benefit from additional security measures, and earn a commission for each successful referral.

- **Data Analytics and Insights:** Aggregate and anonymize data from spam reports and user behavior to generate valuable insights on spam trends and patterns. Offer this data to businesses or researchers for a fee to support their anti-spam efforts or market research.
- **Partnerships with Service Providers:** Collaborate with ISPs, telecom companies, or email service providers to integrate the spam detection app as part of their service packages. Receive a share of the revenue generated from customers who opt for the premium spam protection add-on.
- **Corporate Solutions:** Tailor the app for corporate environments with enterprise-level spam detection and management features. Charge businesses based on the number of users or devices covered by the app.
- **Sponsorships and Brand Collaborations:** Seek sponsorships or collaborations with brands, organizations, or influencers in the cybersecurity space to promote the app and reach a broader audience. Generate revenue from these partnerships and brand endorsements.

It's important to strike a balance between generating revenue and providing value to users. Ensuring that the app remains effective and user-friendly will help drive customer satisfaction and retention, ultimately leading to sustainable revenue growth.

5 Applications of the Project

The spam detection model developed for SMS messages can be applied to solve various other related problems beyond just identifying spam messages. Some of the potential use cases and problems that can be addressed using this model include:

- **Fraud Detection:** The same model can be used to detect fraudulent messages or emails that attempt to deceive users into providing sensitive information or engaging in malicious activities.
- **Phishing Detection:** Phishing emails or messages that impersonate legitimate entities to steal sensitive information can be detected using the spam detection model.
- **Content Moderation:** The model can be applied to moderate user-generated content on online platforms, identifying and filtering out spam or inappropriate messages.
- **Email Filtering:** The model can be adapted to filter spam emails in email clients or servers, helping users manage their inboxes efficiently.

6 Conclusion

In conclusion, the spam detection project has successfully achieved its objective of developing a robust and accurate model for identifying spam messages in SMS communications. Through the implementation of machine learning techniques and natural language processing, the model demonstrated high performance in classifying SMS messages as either spam or non-spam (ham).

Looking forward, the project's success opens avenues for further improvements and expansions. Future work may involve exploring deep learning models to harness complex patterns in text data or conducting research to tackle evolving spamming techniques and challenges. Additionally, ongoing monitoring and updates to the model will ensure its continued effectiveness in combating spam messages in the ever-evolving landscape of communication technologies.

Overall, the spam detection project stands as an essential contribution in mitigating the adverse impact of spam messages on users' communication experiences, enhancing data security, and fostering a safer digital environment for all. Through continuous refinement and adoption in real-world applications, the project can have a meaningful and lasting impact in the field of information security and user protection.

References

- [1] Q. Xu, E. W. Xiang, Q. Yang, J. Du, and J. Zhong, "Sms spam detection using noncontent features," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 44–51, 2012.
- [2] M. Gupta, A. Bakliwal, S. Agarwal, and P. Mehndiratta, "A comparative study of spam sms detection using machine learning classifiers," in *2018 eleventh international conference on contemporary computing (IC3)*, pp. 1–7, IEEE, 2018.
- [3] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *Journal of Big Data*, vol. 2, no. 1, pp. 1–24, 2015.