SMS Spam Detection Using Machine Learning

Dr.M.Sengaliappan[1], C. Deepan Raj[2]

[1]Head of the Department, Department of Computer Applications,

Nehru College of management, Coimbatore, Tamilnadu, India

ncmdrsengaliappan@nehrucolleges.com

[2]Student of II MCA, Department of Computer Applications

Nehru College of management, Coimbatore, Tamilnadu, India

deepanraj0314@gmail.com

Abstract: SMS spam detection is a crucial problem in mobile communication, as unsolicited and potentially harmful messages can lead to privacy invasion, fraud, and disruption of user experience. This study presents an effective approach to identifying spam SMS using machine learning techniques. By leveraging natural language processing (NLP) methods to extract features from text messages, we develop a predictive model that classifies SMS into spam or legitimate categories. The dataset used in this research consists of a large corpus of labelled SMS messages. Various preprocessing steps, such as tokenization, stop word removal, and stemming, are applied to convert raw text into meaningful features. These features are then used to train machine learning models, including Naive Bayes, Support Vector Machines (SVM), Random Forest and like Long Short-Term Memory (LSTM).

Keywords: SVM, Naïve Bayes, Random Forest, Clasification modules, Long Short-Term Memory (LSTM).

I.INTRODUTION:

In the digital age, Short Message Service (SMS) remains a popular communication tool due to its simplicity and ubiquity. However, with the growing use of SMS, there has been an increase in unsolicited and malicious messages, commonly known as spam. SMS spam can range from advertisements and phishing attempts to fraudulent schemes, posing significant threats to privacy and security.

Machine learning-based SMS spam detection involves training algorithms on labeled datasets of spam and non-spam (ham) messages to recognize the patterns and characteristics that distinguish them. By leveraging techniques such as natural language processing (NLP) and statistical analysis, ML models can analyze the content of messages, detect anomalies, and improve their accuracy over time.

In this context, the development and application of machine learning algorithms for SMS spam detection offer a robust, scalable, and adaptive solution to combat spam messages, enhancing user experience and protecting personal information. This introduction explores how machine learning models can be used to address the challenges of SMS spam detection,

discussing various algorithms, features, and evaluation techniques commonly employed in the field.

II.WORKS

SMS spam detection using machine learning works by training algorithms to automatically classify text messages as either spam or legitimate (ham) based on patterns in the message content. The process typically begins with data collection, where a large dataset of labeled SMS messages (spam and ham) is gathered. Next, the text of these messages is pre-processed, involving steps such as tokenization, removing stopwords, and stemming or lemmatization to convert the text into a format suitable for machine learning. Various features, such as word frequency, presence of certain keywords, message length, and special characters, are extracted from the pre-processed text.

These features are then fed into machine learning algorithms like Naive Bayes, Support Vector Machines (SVM), which are trained to recognize patterns in spam messages. During training, the algorithm learns to distinguish between spam and ham based on these patterns. Once trained, the model can predict whether a new, unseen SMS message is spam or not. The model's performance is typically evaluated using metrics like accuracy, precision, recall, and F1 score to ensure its reliability. Over time, the model can improve by incorporating feedback and new data, making it more effective at adapting to evolving spam tactics.

III.MACHINE LEARNING

APPROCHES

SMS spam detection offering unique strengths in terms of performance and complexity. These approaches primarily focus on feature extraction and the application of different algorithms to classify messages as spam or ham. Below are some key machine learning approaches for SMS spam detection:

Support Vector Machines (SVM):

Support Vector Machines (SVM) are a powerful and widely used machine learning algorithm for text classification tasks like SMS spam detection. SVM is particularly effective in high-dimensional spaces, which makes it well-suited for analyzing textual data where features can be sparse (i.e., many different words or word combinations).

Naive Bayes:

Naive Bayes is a popular and efficient machine learning algorithm for text classification tasks, including SMS spam detection. It is based on Bayes' Theorem and is called "naive" because it assumes that the features (words in an SMS) are independent of each other, which is rarely true in practice.

Naive Bayes calculates the probability of a message being spam or ham (non-spam) based on the words or features present in the message. It does so by using the **Bayes Theorem**, which is defined as:

$$P(\text{Class}|\text{Message})=P(\text{Message})/P(\text{Message}|\text{Class})\cdot P(\text{Class})$$

	target	text
1418	ham	Lmao. Take a pic and send it to me.
2338	ham	Alright, see you in a bit
88	ham	I'm really not up to it still tonight babe
3735	ham	Hows the street where the end of library walk is?
3859	ham	Yep. I do like the pink furniture tho.

Figure 1. Naïve Bayes

Random Forest:

Random Forest is an ensemble learning method that combines the predictions of multiple decision trees to improve classification accuracy and control overfitting. It is widely used in various classification tasks, including SMS spam detection, due to its robustness, flexibility, and ability to handle large datasets with high dimensionality.

IV. METHODOLOGY

The methodology for SMS spam detection involves a systematic approach that encompasses data collection, preprocessing, feature extraction, model selection, training, evaluation, and deployment.

Data Collection: Data collection is a crucial initial step in developing an SMS spam detection system. The quality and diversity of the dataset directly influence the model's performance.

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
2464	ham	They will pick up and drop in car.so no problem..	NaN	NaN	NaN
1248	ham	HI HUN! IM NOT COMIN 2NITE-TELL EVERY1 IM SORR...	NaN	NaN	NaN
1413	spam	Dear U've been invited to XCHAT. This is our f...	NaN	NaN	NaN
2995	ham	They released vday shirts and when u put it on...	NaN	NaN	NaN
4458	spam	Welcome to UK-mobile-date this msg is FREE giv...	NaN	NaN	NaN

Figure 2. Data Collection

Data Preprocessing:

Data preprocessing prepares raw text data for effective feature extraction and model training. Here's a concise overview of the key preprocessing steps involved:

1. Text Cleaning
2. Stop Word Removal
3. Tokenization
4. Stemming
5. Handling Imbalanced Data

	target	text	num_characters	num_words	num_sentences	transformed_text
0	0	Go until jurong point, crazy.. Available only ...	111	24	2	go jurong point crazi avail bugi n great world...
1	0	Ok lar... Joking wif u oni...	29	8	2	ok lar joke wif u oni
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2	free entri 2 wkli comp win fa cup final tkt 21...
3	0	U dun say so early hor... U c already then say...	49	13	1	u dun say earli hor u c alreadi say
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1	nah think goe usf live around though

Figure 3. Data Preprocessing

TRAINING DATA:

Training data is a critical component of the machine learning process, serving as the foundation upon which models learn to make predictions. It consists of a subset of data used to train a machine learning algorithm, allowing it to identify patterns and relationships in the data.

	Algorithm	Accuracy	Precision
1	KN	0.900387	1.000000
2	NB	0.959381	1.000000
8	ETC	0.977756	0.991453
5	RF	0.970019	0.990826
0	SVC	0.972921	0.974138
6	AdaBoost	0.962282	0.954128
10	xgb	0.971954	0.950413
4	LR	0.951644	0.940000
9	GBDT	0.951644	0.931373
7	BgC	0.957447	0.861538
3	DT	0.935203	0.838095

Figure 4. Sample Training Data

MODIFY DATA:

Data modification in machine learning involves transforming and preparing raw data to enhance its quality, structure, and suitability for analysis and modeling. Techniques such as data cleaning, transformation, feature engineering, and outlier treatment help improve data quality.

	Algorithm	variable	value
0	ETC	Accuracy	0.977756
1	SVC	Accuracy	0.972921
2	xgb	Accuracy	0.971954
3	RF	Accuracy	0.970019
4	AdaBoost	Accuracy	0.962282
5	NB	Accuracy	0.959381
6	BgC	Accuracy	0.957447
7	LR	Accuracy	0.951644
8	GBDT	Accuracy	0.951644
9	DT	Accuracy	0.935203
10	KN	Accuracy	0.900387

Figure 5. Modify Training Data

```
Accuracy -  0.8665377176015474
Precision -  0.0
For  KN
Accuracy -  0.9284332688588007
Precision -  0.7711864406779662
For  NB
Accuracy -  0.9400386847195358
Precision -  1.0
For  DT
Accuracy -  0.9439071566731141
Precision -  0.8773584905660378
For  LR
Accuracy -  0.9613152804642167
Precision -  0.9711538461538461
For  RF
Accuracy -  0.9748549323017408
Precision -  0.9827586206896551
```

Figure 6. Model Evaluation

V.Model Evaluation and Performance

Evaluating and improving the performance of an SMS spam detection model using machine learning involves a combination of selecting the right algorithms, applying effective pre-processing techniques, and using appropriate metrics for model assessment.Algorithms are

- **Naive Bayes**: Works well for text data (multinomial Naive Bayes is common for spam detection).
- **Logistic Regression**: Simple yet effective for binary classification.
- **Support Vector Machines (SVM)**: Can handle high-dimensional text data.
- **Random Forest**: Captures non-linear relationships and is robust.

Accuracy:

Accuracy can be misleading if the data is imbalanced.

Accuracy=TP+TN/TP+TN+FP+FN

Precision:

The percentage of true spam messages out of all messages classified as spam (focus on minimizing false positives).

Precision=TP/FP+TP

Recall:

Measures how many actual spam messages were correctly identified .

Recall=True Positives + False Negatives / True Positives

F1 Score:

Harmonic mean of precision and recall, providing a balanced metric for uneven class distribution.

VI.RESULTS:

The model for spam email detection using machine learning performs well, achieving high accuracy and F1-scores. With a precision of over 90%, the model minimizes the risk of false positives, ensuring that important emails are not flagged as spam.

Model	Accuracy	Precision	Recall	F1 Score	ROC - AUC
Naïve Bays	98.2	96.0	89.0	92.3	97.0
Logistic Regression	98.7	97.2	91.5	94.3	98.1
SVM	99.0	98.3	92.7	95.6	98.5

VI.CONCLUSION:

Machine learning has proven to be highly effective for SMS spam detection, offering robust solutions for filtering spam messages while maintaining user experience. By leveraging various models such as **Naive Bayes**, **Logistic Regression**, and **Support Vector Machines (SVM)**, the task of distinguishing between legitimate (ham) and spam messages can be accomplished with high accuracy and reliability. Machine learning models for SMS spam detection can achieve high levels of accuracy, often exceeding **98%**, with the best models achieving strong precision and recall balances. By continuously updating models with new data and using appropriate preprocessing techniques, organizations can maintain effective and efficient spam filters to enhance user experience and security.

As spam tactics evolve, maintaining and refining models will be important to ensure they adapt to new patterns and continue to provide reliable spam detection.

References:

[1] https://www.who.int/hrh/links/en/

[2] https://en.wikipedia.org/wiki/Machine learning

[3] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, 2017, pp.204-207, doi: 10.1109/ISCC.2017.8024530.

[4] S. Dhar, K. Roy, T. Dey, P. Datta and A. Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777531.

[5] C. Raju, E. Philipsy, S. Chacko, L. Padma Suresh and S. Deepa Rajan, "A Survey on Predicting Heart Disease using Data Mining Techniques," 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, 2018, pp. 253-255, doi: 10.1109/ICEDSS.2018.8544333.

[6] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid,S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, “International application of a new probability algorithm for the diagnosis of coronary artery disease,” The American journal of cardiology, vol. 64, no. 5, pp. 304–310, 1989.

[7] B. Edmonds, “Using localised ’gossip’ to structure distributed learning,” 2005.

[8] Fsdfsdf BayuAdhi Tama,1 Afriyan Firdaus,2 Rodiyatul FS, “Detection of Type 2 Diabetes Mellitus with Data Mining Approach Using Support Vector Machine”, Vol. 11, issue 3, pp. 12-23, 2008.