International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 10 Issue 3, May-Jun 2026 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Natural Language Processing for Span Detection

Matthew N. O. Sadiku[1], Paul A. Adekunte[2], Janet O. Sadiku[3]

[1]Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA
[2]International Institute of Professional Security, Lagos, Nigeria
[3]Juliana King University, Houston, TX, USA

ABSTRACT

Email (or electronic mail) is a popular communication method used in both internal and internet networks for information exchange. While email has significant benefits, its misuse can have negative impacts if not used wisely. Spam (meaning Stupid Pointless Annoying Messages) comes in various types, including advertisements, phishing, malware viruses, scams, and others. Spam is the pervasive nature of unsolicited and often malicious messages. The impact of spam extends beyond annoyance, as it poses risks to personal information security and hampers productivity. In its various insidious forms, spam continues to plague digital communication channels, ranging from email inboxes to social media platforms. Spam detection has evolved significantly from rudimentary rule-based systems to sophisticated approaches leveraging natural language processing (NLP) and machine learning. NLP has become the cornerstone of modern defense systems because NLP approach provides better performance in spam detection compared to other methods. This paper explores the pivotal role of NLP in spam detection.

KEYWORDS: *natural language processing (NLP), computational linguistics, spam detection, emails, artificial intelligence, machine learning.*

How to cite this paper: Matthew N. O. Sadiku | Paul A. Adekunte | Janet O. Sadiku "Natural Language Processing for Span Detection" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-10 | Issue-3, June 2026, pp.349-356, URL: www.ijtsrd.com/papers/ijtsrd102030.pdf

IJTSRD102030

INTRODUCTION

It is rare to find people using postal services to send letters, as physical letters are no longer the primary means of communication between individuals. Letters have transformed into a digital format known as email. The evolution of digital communication has been shadowed by the persistent growth of unsolicited and malicious content/email, commonly known as spam. Spam, from unsolicited emails to fraudulent messages, poses a significant threat to digital communication, impacting productivity, security, and user experience. From email spam to SMS phishing (smishing), these deceptive communications aim to extract sensitive information or propagate malware. Email remains one of the most exploited vectors for cyber threats, including unsolicited commercial spam and targeted phishing attacks. The relentless evolution of spam tactics necessitates sophisticated detection mechanisms. Natural language processing (NLP), a branch of artificial intelligence (AI) focused on enabling computers to understand, interpret, and generate human language, has emerged as a critical tool in this ongoing battle. By analyzing the linguistic characteristics of messages, NLP techniques can effectively distinguish legitimate communications from malicious ones [1].

Spam has gotten out of control and is now a serious problem that affects information security and user experience due to the exponential rise of email communication. Spams go through email billions of times on an average day, per cybersecurity report. Such reports could lead into financial scams, identity theft, and wasted resources. It is therefore critical as part of the natural language processing and the general picture of cybersecurity that spam should be detected. Traditional mechanisms such as rule-based filtering, keyword matching, and blacklist/whitelist-type conventional spam filtering systems are no longer sufficient, as spammers have developed ways of circumventing such measures. The conventional methods also suffer from many false positives, rigidity, and failure to infer semantic meaning

embedded in email text. To overcome these limitations, researchers have devoted their efforts toward building intelligent and adaptive spam detection mechanisms based on machine learning (ML) and natural language processing (NLP) techniques [2].

FUNDAMENTALS OF NLP

Natural language processing is a subfield of artificial intelligence that empowers computers to understand, interpret, and generate human language. It is a technique where machine can become more human and thereby making human to communicate with the machine easily. NLP seeks to make software intelligent enough to process a natural language as humans. For example, imagine a machine that takes instructions by voice.

NLP analysis generally consists of the following three levels [3]:

- *Syntax,* the study of sentence structure. Syntax deals with the formation of a sentence from individual words. Syntax alone suggests the proper interpretation of "Jimmy loves Lucy."
- *Semantics,* the study of context-independent meaning. This derives the meaning of a sentence based on the meanings of the words/phrases. For example, semantics determines whether the word "bank" refers to a river bank or to a financial institution.
- *Pragmatics,* the study of context-dependent meaning. Pragmatics deals with how meaning changes in the presence of a specific context and how the contexts affect the meaning of the sentences. This level is concerned with the purposeful use of language in situations.

As a foundational pillar of modern artificial intelligence, NLP encompasses a wide array of tasks, including speech recognition, text classification, natural language understanding (NLU), and natural language generation (NLG). NLP encompasses a wide range of tasks, such as information retrieval (IR), named entity recognition (NER), relation extraction, text classification, topic modeling, semantic textual similarity, machine translation, and question answering (QA). Figure 1 shows how NLP transforms raw acoustic data into meaningful interactions [4], while Figure 2 shows different components of NLP [5].

Recently, large language models (LLMs) have shown their ability in learning universal language representations, text understanding and generation. LLMs refer to a model with a large number of parameters, vast training data, and substantial compute, enabling it to capture complex language patterns. In LLM-based NLP, pre-processing is followed by prompt engineering, which guides LLMs to produce outputs that align with extraction requirements during inference without altering the model's parameters. Models like GPT are pushing the boundaries of language understanding, enabling nuanced and context-aware applications. The GPT (Generative Pretrained Transformer) is a large-scale language model developed by OpenAI that consists of multiple layers of transformer blocks, each with a self-attention mechanism and a forward neural network [6]. GPT-based systems can summarize complex reports or generate creative content like essays, making them versatile in both academic and professional environments. ChatGPT uses NLP techniques to understand prompts. When you enter a prompt, the chatbot comprehends it and provides relevant replies. Figure 3 shows the word cloud for NLP [7]. Figure 4 shows how NLP works [8], while Figure 5 depicts some applications of NLP [9].

NLP IN SPAM DETECTION

Spam emails refer to unsolicited messages sent in large quantities to numerous recipients. They can take various forms, including phishing attempts, promotional offers, and fraudulent schemes. The continuous arms race between spammers and detection systems necessitates increasingly advanced methodologies. Spammers constantly innovate, adopting new vocabulary, topics, and obfuscation techniques to bypass existing filters. A model trained on past spam data may quickly become outdated as new spam campaigns emerge. For instance, the language used in phishing emails today differs significantly from that of a decade ago. This dynamic environment necessitates continuous model retraining and adaptive learning. Traditional spam detection mechanisms were predominantly rule-based systems. They often rely on keyword matching and rule-based filtering and have proven increasingly inadequate against the nuanced and adaptive strategies employed by spammers. This is where natural language processing (NLP) emerges as a pivotal technology, offering a robust and dynamic approach to identifying and mitigating unwanted messages. By enabling machines to understand, interpret, and generate human language, NLP significantly enhances the accuracy, efficiency, and adaptability of spam detection systems [1]. With the power of natural language processing (NLP) and machine learning (ML), we can build a smart spam detector. Figure 6 depicts a spam detection [10], while Figure 7 shows how NLP and ML are used in spam detection [10].

The development of email is not without issues, including the possibility of spam emails entering

someone's account. This causes users difficulty in distinguishing whether an email has the potential to be dangerous or just junk. Spam emails have become an incessant nuisance in our digital lives, flooding our inboxes with unwanted messages that range from promotional offers to phishing scams and fraudulent schemes. Spam can be distinguished from non-spam emails, mainly by observing the subject and content of the email. The content or core of the email sent by spammers is also an important indicator in differentiating spam from other emails. Messages can be classified as spam or ham by applying NLP methods such as tokenizing, part-of-speech tagging, stemming, and chunking. The NLP approach has proven to be one of the effective methods in detecting email spam [11].

APPLICATIONS OF NLP SPAM DETECTION

Spamming is rapidly proliferating across various digital communication channels, with email being the most common medium for spam distribution. Spam detection using natural language processing and machine learning is a powerful way to filter unwanted messages and enhance user safety. Common applications of NLP in spam detection include the following [2,12]:

- *Email Communication:* Email communication is the backbone of personal and professional digital interaction worldwide, with an estimated 347 billion emails sent daily as of 2023. Every form of electronic communication faces the curse of spam, whether it be phishing, online fraud, or malware dissemination. Email was, and continues to be, the most popular mode of communication because it effectively provides both official and personal communication channels. This critical channel is persistently exploited by malicious actors for spam campaigns and phishing attacks. Given the growing risks spam emails present to cybersecurity and communication, developing reliable classification methods is critical. Incoming email text can be classified into three categories — legitimate (ham), spam, and phishing. Figure 8 shows a typical distribution of ham and spam messages [10]. Spam refers to unsolicited bulk email sent for commercial or fraudulent purposes, while phishing is a more targeted form of social engineering designed to deceive recipients into disclosing sensitive information such as passwords, financial credentials, or personal identification data. Traditional defenses against email threats — including blacklists, rule-based filters, and IP reputation systems — have become increasingly inadequate against modern attacks that employ varied vocabulary, obfuscated URLs, and personalized content. The application of natural language processing (NLP) and machine learning techniques offers a more adaptive and semantically aware detection paradigm. Two key technologies for detecting and filtering spam emails are machine learning and knowledge engineering.
- *Spam Filter Interpretability*: Existing systems tend to either under-or overweight accuracy without considering the interpretable outcomes that would be usable for end-users and corporations during their decision-making. Besides reporting competitive accuracy, spam filter provides transparency, thus facilitating easy implementation in real setting. NLP needs modification to make it machine-interpretable, and requires multiple stages of processing for feature extraction.
- *Classification:* Spam detection is often accomplished by the utilization of learning-based classifiers. In the context of learning-based classification, the approach to detection operates under the assumption that spam emails possess distinct properties that can be used to identify them as valid emails. Classification problems can be broadly split into two categories: binary classification problems and multi-class classification problems. Binary classification means there are only two possible label classes, e.g. a patient's condition is cancerous or it is not. Multi-class classification refers to cases where there are more than two label classes. An example of this is classifying the sentiment of a movie review into positive, negative, or neutral. There are many types of NLP problems, and one of the most common types is the *classification of strings*. Examples of this include the classification of movies/news articles into different genres, and the automated classification of emails into spam or not spam. A classifier separates emails as spam or not spam. Figure 9 shows a typical classifier [10].

BENEFITS

One of the primary benefits of NLP in spam detection lies in its contextual understanding capabilities. Unlike rudimentary filters that merely scan for predefined keywords, NLP-powered systems delve deeper into the semantic and syntactic structure of messages. This ability to interpret language despite deliberate obfuscation provides a significant advantage over rule-based systems, which are easily circumvented by minor alterations to known spam patterns. NLP also excels in detecting sophisticated

and obfuscated spam. Another critical advantage is the reduction of false positives and false negatives. NLP-based spam detection systems have proven highly effective in identifying and filtering spam across social media platforms and email systems. Other benefits of NLP in spam detection include the following [1,13]:

- *Enhanced Security:* By analyzing linguistic patterns, sentiment, and the overall coherence of a message, NLP systems can more accurately differentiate between genuine communication and deceptive content. For example, an NLP model can distinguish between a legitimate email from a bank regarding an account update and a phishing attempt that mimics such a message, thereby minimizing the disruption caused by false positives and enhancing security by catching more sophisticated threats.
- *Adaptability:* NLP-driven spam detection systems offer superior adaptability and continuous learning. The landscape of spam is constantly evolving, with new techniques and themes emerging regularly. The iterative learning process allows the systems to stay current with emerging spam trends and adversarial attacks, such as those involving generative AI to create highly convincing spam messages. This dynamic adaptability ensures that NLP-based filters remain effective against the ever-changing tactics of spammers, providing a resilient defense mechanism in the ongoing battle against unwanted digital intrusions.
- *Improved Accuracy:* NLP-based spam filters demonstrate higher accuracy in distinguishing between spam and legitimate emails, resulting in fewer false positives and false negatives.
- *Reduced False Positives/Negatives*: By leveraging the contextual understanding and semantic analysis capabilities of NLP, false positives (legitimate emails marked as spam) and false negatives (spam emails marked as legitimate) can be minimized.

CHALLENGES

In spite of the significant progress, spam detection remains an arms race and its application is fraught with significant challenges, ranging from the inherent complexities of human language to the deliberate subversion tactics employed by adversarial actors. Detecting spam emails poses several challenges due to their evolving nature and the sheer volume of messages being sent. Spammers continuously adapt their strategies, leading to new challenges. They constantly refine their techniques, making it difficult for traditional spam filters to keep up. Challenges like evolving spam tactics and multilingual detection persist. Other challenges of NLP in spam detection include the following [1,13]:

- *Human Language*: One of the primary hurdles for NLP-based spam filters is the fluid and nuanced nature of human language. Traditional filters often relied on keyword-based heuristics, but modern spammers exploit linguistic features that are difficult for machines to interpret accurately. The rapid evolution of slang, professional jargon, and cultural idioms creates a "moving target" for NLP models.
- *Ethical Concerns*: As NLP-based spam detection becomes more powerful, ethical considerations surrounding privacy and data protection must be carefully addressed to strike the right balance between security and user experience.
- *Data Imbalance:* In real-world scenarios, the volume of legitimate (ham) messages significantly outweighs that of spam messages. See Figure 8, for example. The cost of a false positive (flagging a critical legitimate email as spam) is often far higher than the cost of a false negative (letting a spam message through). This imbalance forces developers to tune models with high precision, which inherently limits their recall and allows more sophisticated spam to leak through. The data imbalance can lead to models that are biased towards the majority class (ham), resulting in higher false negative rates (spam messages being classified as legitimate). Addressing this requires techniques such as oversampling the minority class, undersampling the majority class, or using specialized algorithms designed for imbalanced datasets.
- *Multilingual Spam:* Spam is no longer limited to text; it often incorporates images, videos, and audio. Detecting multimodal spam requires integrating NLP with computer vision and audio processing techniques. The global nature of communication means spam can originate in various languages and scripts. Developing robust spam detection systems that can effectively handle multilingual spam is a complex challenge, requiring models capable of processing diverse linguistic features and cultural contexts.
- *Adversarial Arms Race*: The most significant challenge in NLP-based spam detection is the presence of an active, intelligent adversary. Unlike other NLP tasks like translation or sentiment analysis, where the data is generally cooperative, spam detection involves a constant

"cat-and-mouse" game. Malicious actors employ adversarial techniques to craft spam messages that can bypass detection systems. These attacks involve subtle modifications to text that are imperceptible to humans but can trick machine learning models into misclassifying spam as legitimate. Defenses against such attacks involve robust training methods and adversarial training.

- *Adversarial Training:* To enhance model robustness against adversarial attacks, models can be explicitly trained on adversarial examples. This process exposes the model to perturbed inputs during training, making it more resilient to similar attacks in deployment.
- *Online Learning and Adaptive Models*: To combat concept drift, spam filters need to be continuously updated. Online learning approaches allow models to adapt in real-time or near real-time to new patterns of spam, ensuring they remain effective against emerging threats. This often involves retraining models periodically or using incremental learning techniques.
- *Linguistic Variations:* Language-specific challenges and linguistic variations across different regions and cultures can pose difficulties in accurately detecting spam emails.

FUTURE OF NLP SPAM DETECTION

In the battle against spam emails, natural language processing (NLP) has emerged as a powerful weapon. The field of NLP continues to evolve, paving the way for exciting advancements in spam detection. The future of spam detection will likely involve a multi-layered approach, combining advanced NLP with other AI techniques. As generative AI continues to lower the barrier for creating sophisticated spam, the future of defense lies in multi-layered strategies that combine advanced semantic analysis with behavioral heuristics and robust adversarial training.

The continuous evolution of NLP, particularly with advancements in LLMs, will be crucial in developing adaptive and resilient spam detection systems. In future technological developments, the NLP approach has the potential for further development to address new challenges and enhance information security in emails. In its evolution, NLP techniques can be expanded by applying deep learning methods such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) to improve accuracy and effectiveness in detecting email spam. Advancements in contextual understanding and semantic analysis will enable NLP systems to better comprehend the nuances and subtleties of human language, improving spam detection capabilities. Although NLP faces challenges such as linguistic variations and adversarial attacks, the future looks promising with advancements in contextual understanding and integration with other AI techniques. As NLP continues to evolve, it holds the potential to combat spam effectively [13].

CONCLUSION

Natural language processing has revolutionized spam detection, transforming it from a rule-based endeavor into a proactive, intelligent system driven by AI. It has changed spam detection by moving beyond simplistic keyword matching to a more profound understanding of language. The dynamic nature of spam, characterized by concept drift and sophisticated adversarial attacks, ensures that the fight against unwanted messages remains an ongoing challenge. The capabilities of NLP in contextual comprehension, handling obfuscation, reducing classification errors, and continuous adaptation make it an indispensable tool in safeguarding digital communication. However, the battle against spam is dynamic, with new challenges emerging from adversarial attacks and the proliferation of AI-generated content.

The integration of NLP into spam detection has undoubtedly improved our ability to secure digital communication, yet the challenges remain formidable. The NLP approach has the potential to address new challenges and enhance information security in emails. It excels in providing more accurate results in email spam detection compared to conventional approaches. More information about the integration of NLP in spam detection can be found in the following related journals:

- Natural Language Processing Journal
- Journal of Emerging Technologies and Innovative Research

REFERENCES

[1] https://manus.im

[2] S. Patha et al., "Spam detection for emails using natural language processing and explainable machine learning," *International Journal of Research Publication and Reviews,* vol. 6, no. 8, August 2025, pp 5444-5451.

[3] J. Hirschberg, B. W. Ballard, and D. Hindle, "Natural language processing," *AT&T Technical Journal*, Jan./Feb. 1988, vol. 67, no. 1, 1988.

[4] A. Jain, "What is the role of NLP in voice assistants?" August 2024, https://www.analyticsinsight.net/nlp/what-is-the-role-of-nlp-in-voice-assistants#google_vignette

[5] "How Google uses NLP to improve SERPs, featured Snippets & UX," https://digitalguider.com/blog/what-is-google-nlp/

[6] X. Jiang et al., "Applications of natural language processing and large language models in materials discovery," *NPJ Computational Materials*, vol. 11, no.79, 2025.

[7] P. Pagade, "NLP (natural language processing) for spam detection," November 2016, https://www.linkedin.com/pulse/nlp-natural-language-processing-spam-detection-prasad-pagade

[8] "What is NLP? How it works, benefits, challenges, examples," June 2025, https://www.shaip.com/blog/what-is-nlp-how-it-works-benefits-challenges-examples/

[9] A. Arif, "NLP in finance: Examining the impact of natural language processing in financial and banking services," July 2023, https://www.johnsnowlabs.com/examining-the-impact-of-nlp-in-financial-services/

[10] "Using natural language processing for spam detection in emails," September 2020, https://medium.datadriveninvestor.com/using-natural-language-processing-for-spam-detection-in-emails-281a7c22ddbc

[11] G. S. M. Diyasa et al., "Implementation of natural language processing for spam email detection in outcome based education (OBE) application," *International Journal of Entrepreneurship and Business Development,* vol. 6, no. 6, November 2023.

[12] Shweta and S. Daniel, "AI-driven spam and phishing email detection using natural language processing," *International Journal of Versatile Research and Analysis*, vol. 5, no. 5, May 2026.

[13] D. Thakran, "Exploring the significance of NLP in effective spam detection," July 2023, https://astconsulting.in/artificial-intelligence/nlp-natural-language-processing/exploring-the-significance-of-nlp-in-effective-spam-detection

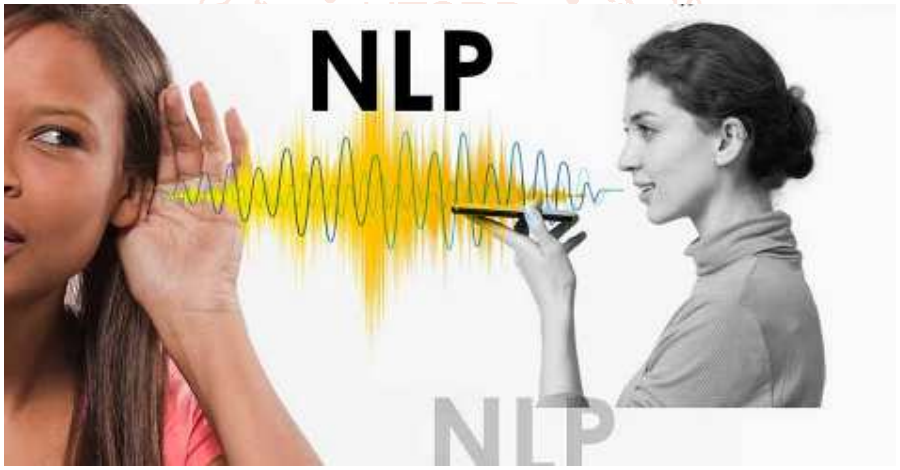

Figure 1 A representation of NLP [4].

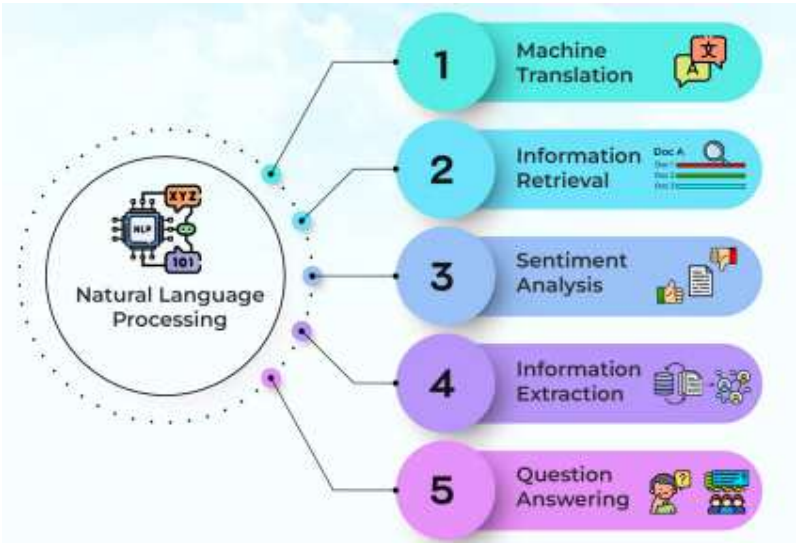

Figure 2 Different components of NLP [5].

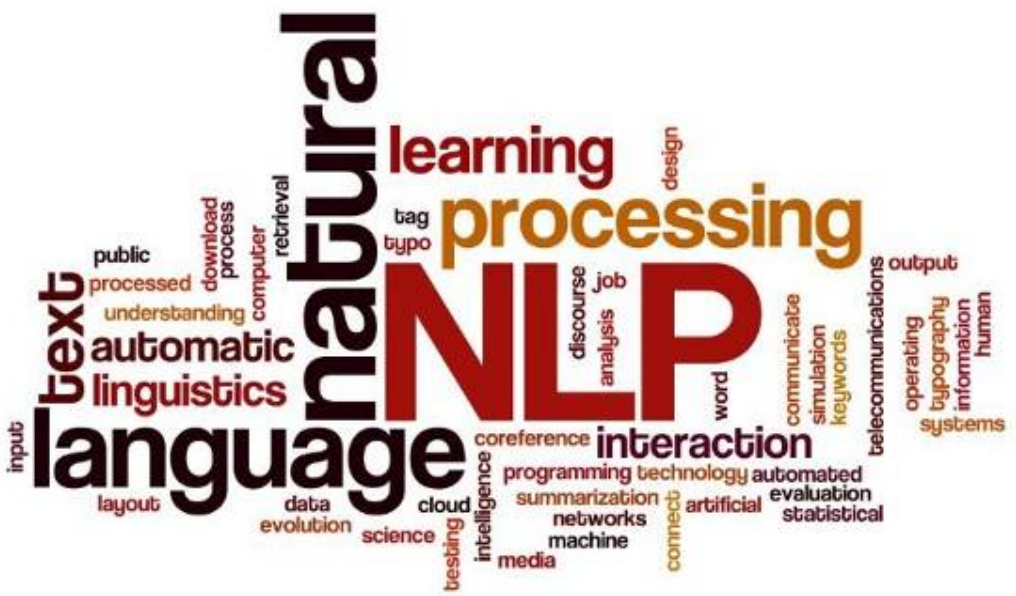

Figure 3 The word cloud for NLP [7].

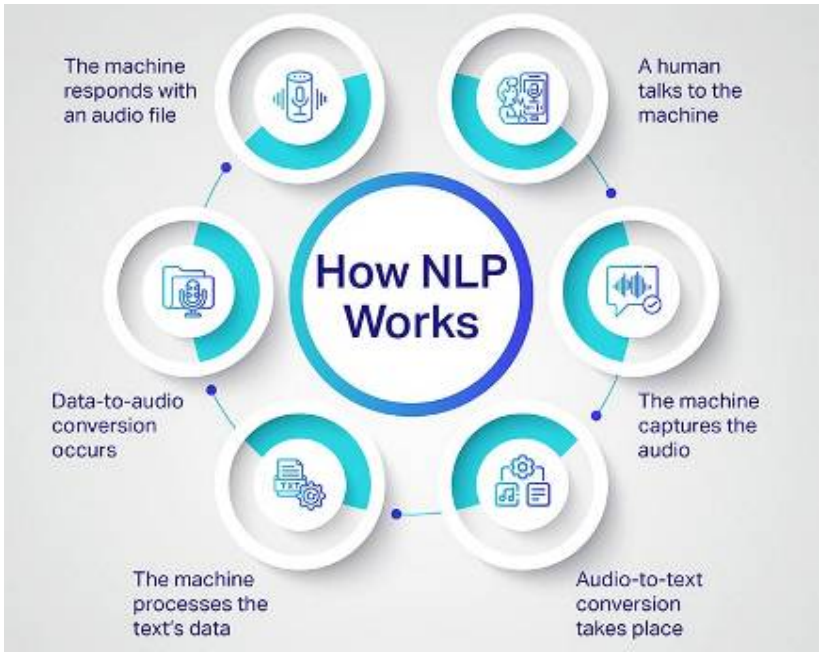

Figure 4 How NLP works [8].

Figure 5 Some applications of NLP [9].

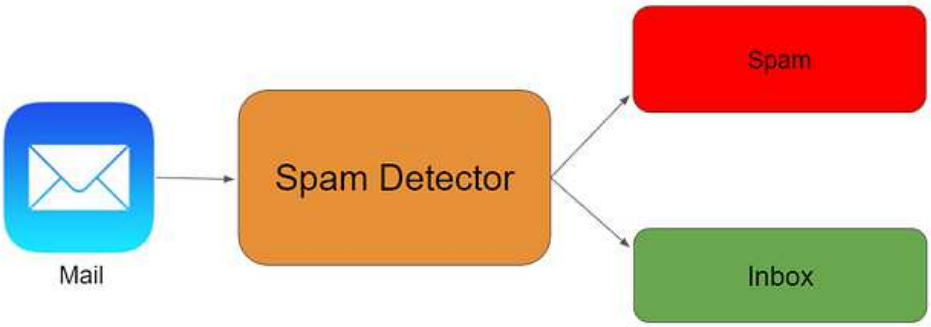

Figure 6 Spam detection [10].

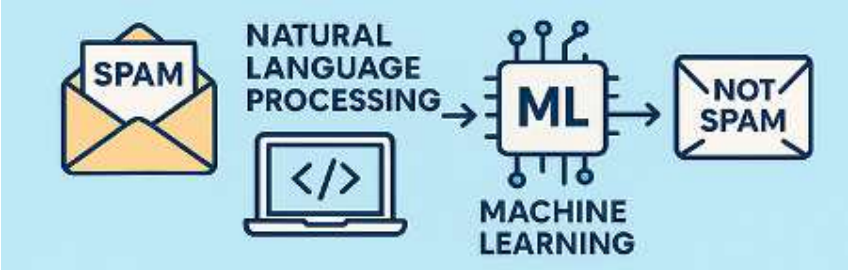

Figure 7 How NLP and ML are used in spam detection [10].

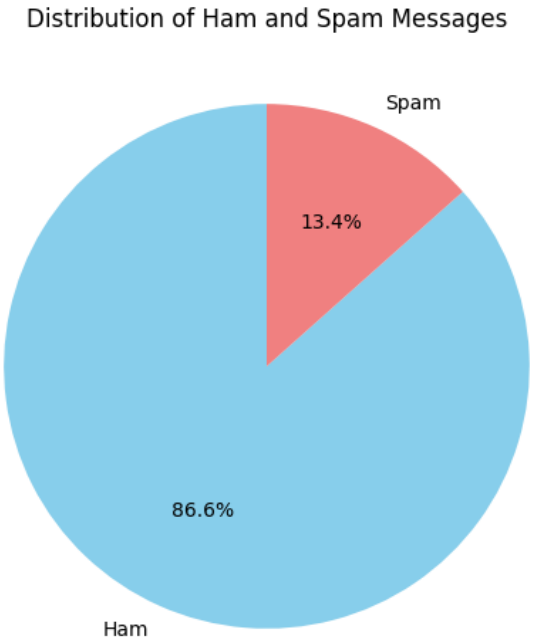

Figure 8 A typical distribution of ham and spam messages [10].

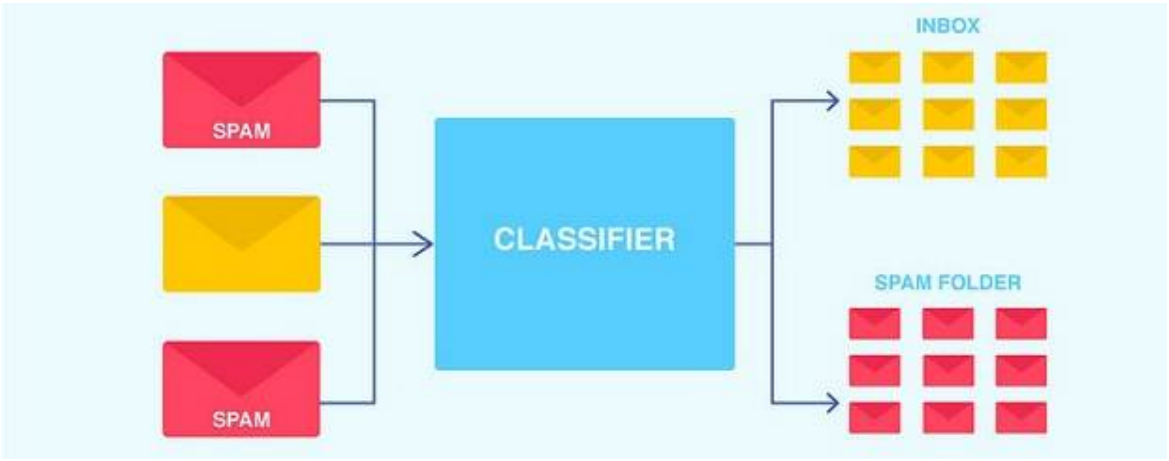

Figure 9 A typical classifier [10].