Journal of Policy Research, 9(4), 152-162
https://jprpk.com
https://doi.org/10.61506/02.00136

A Review to Unfold the Role of Machine Learning Algorithms in Natural Language Processing

Arslan Ali Raza[1], Khansa Qamar[2], Shirin Fatima[2], Komal Fatima[2*], Ali Arslan[2], Hira Tahseen[2], Hazaima Aslam[2], Uzma Parveen[2], Ayesha Asghar[2]

Abstract
This research review has been written to explore the role of machine learning algorithms in natural language processing (NLP), a computer field that offers human-like comprehension of virtual text. The main motive of the article is to reveal the participation of machine learning (ML) in NLP and its scope in coherent fields. In this study, the researchers have used a systematic literature review approach to explore the role of ML algorithms in NLP. Highlighting the techniques of ML algorithms as supervised, unsupervised, semi-supervised, and reinforcement methods this study discloses the connectivity with morphological, semantic, syntactic, pragmatic, and discourse analysis in NLP. The article describes that ML/NLP has immense applications in different fields where several tools of ML/NLP are utilized. The researchers have gone through the study of the use of chatbots, text summarization tools, web scrapping, sentimental analysis in the social media stock market, medical field disease detection, and fraud detection. ML's role is detained in current studies along with the contributions to the progressions. A distinct comparison has been made between the past present and future of ML in NLP. ML algorithms and applications with examples are disclosed in the classification of logistic regression, SVM, Naive Bayes, K-Nearest Neighbor, and decision tree. Data distortion, interpretation of the research findings, and contextual ambiguity are visibly determined as challenges and discussed in the study. This writing opens up to the recent advancements and advantages in multiple areas correspondingly. The discussions and findings expose the role of ML in opening the gateway to the profound revolutionary search engines, algorithms, and multiple techniques through the development of technology and its evolution to adapt to new language differences, slang, and changes in language use.
Keywords: Algorithm, Models, Machine Learning, AI, Networks, Computer

1. Introduction

Natural language processing (NLP) is a field within computer science that works to provide computer systems with a human-like understanding of spoken language and digital text (Kili, 2023) . As to dealing with human languages in spoken or text forms of data, it mixes machine learning, statistical modeling, deep learning standards, and computational linguistics. This process aims to figure out the message, purpose, and beliefs of the respondent. The reason behind this process is that human language is kind of difficult to understand, as in our daily lives, we use slang, sarcasm, and sometimes abbreviations to convey our emotions or feelings. The computer system is unable to understand human language due to its complex structure. So, natural language processing (NLP) helps computer system to understand natural human language (Bansal, 2023). In the past, it was difficult for the computer to understand complex human language, but with the development of natural language processing (NLP) and machine learning (ML), the computer can now understand spoken and digital text. Natural language processing is used by computer programs to translate digital text, respond accordingly and fluently to spoken commands, and instantly summarize huge amounts of data. Voice-guided GPS automobile systems, voice recognition software, Chabot's for customer support, and personal digital assistants such as Alexa, Siri, etc. all frequently use NLP. Moreover, it is a key component of enterprise-based solutions, which help improve corporate operations, enhance employee efficiency, and support major business procedures. Currently, machine learning (ML) and natural language processing (NLP) are the two important and distinct functions of artificial intelligence (Khurana, Koli, Khatter, & Singh, 2022). The primary objective of natural language processing is to interpret human language and how it is programmed in a machine, semantically, syntactically, and pragmatically. Natural language processing (NLP) can be employed in multiple fields such as Information Retrieval, Information Extraction, Summarization, Question Answering, Email Spam detection, Auto-correct, Auto-fill, Chatbot systems, Speech Recognition, Semantic Analysis, Dialogue System, and medicine. Deep Learning, a subfield of Artificial Neural Networks, is widely adopted and has produced good results in natural language processing research. Techniques like Restricted Boltzmann machines, Auto encoders, Deep Neural networks (DNN), Recurrent Neural networks (RNN), and Convolution Neural networks (CNN) are used to achieve accuracy in various applications, making them the most significant aspect of machine learning techniques in the field.

2. Machine Learning Techniques

Machine Learning algorithms are mainly divided into four learning methods in NLP; supervised, unsupervised, semi-supervised, and reinforcement. Natural language processing (NLP) employs both supervised and unsupervised

[1] Department of Computer Science, COMSATS University Islamabad, Vehari Punjab Pakistan, arslan@cuivehari.edu.pk
[2*] Department of Humanities, COMSATS University Islamabad, Vehari Punjab Pakistan, fatimakomal57@gmail.com

machine learning methods. While supervised learning models demand data with labels, unsupervised learning models employ unlabeled data. In natural language processing (NLP), neural networks (NN), such as convolutional neural networks (CNN) and recurrent neural networks (RNN), are frequently employed for task cycles like language translation and text categorization applications as sentiment analysis. Natural language processing also includes unsupervised learning (NLP). This kind of machine learning makes use of topic modeling and clustering (ML). Topic modeling can be used to determine the underlying subjects within a collection of documents, while clustering algorithms can be used to group related topics together. Because both methods can analyze large volumes of text, comparable publications can also be put together in clustering. Neural network (NN) such as in natural language processing (NLP). The technique of locating, obtaining, and classifying the opinions expressed in a text is known as sentiment analysis. It can be applied to market inquiry, service to customers, and surveillance of the media. Understanding whether a specific piece of writing has a positive, negative, or impartial tone is the primary objective of sentiment analysis. Opinion mining or sentiment classification are common terms used to describe this. Natural Language Understanding (NLU) and Natural Language Generation (NLG) are the two components of Natural Language Processing (Goyal, 2023). NLU is about linguistics that deals with all the functions of linguistics, NLG functions to generate natural language.

Different phases involved in the processing and analysis of human language include morphological analysis, syntactic analysis, semantic analysis, discourse analysis, and pragmatic analysis as shown clearly in Figure 1 (Goyal, 2023).

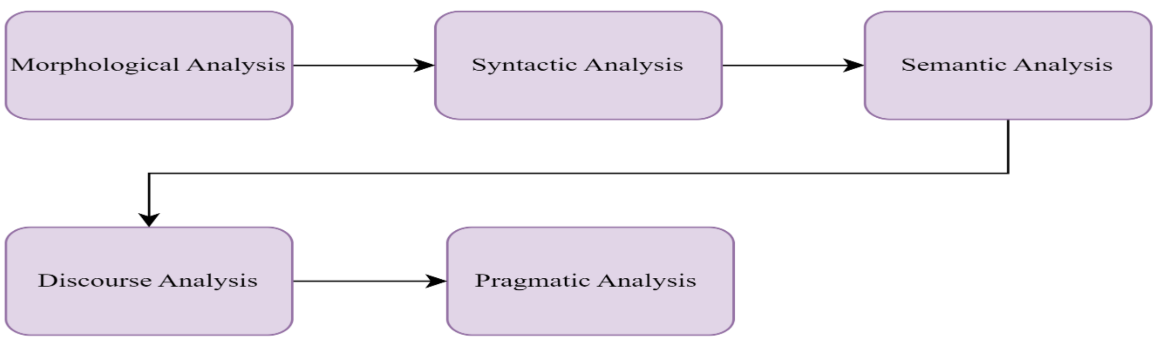

Figure 01: Analysis of Human Language in NLP

This analysis process determines the computer's capability of processing and comprehending natural human language. In the morphological phase of NLP, affixes are removed through tokenization and stemming and moved to the syntactic analysis phase, the natural language process (NLP) categories words, and the syntactic parser checks grammatical rules through recurrent neural network (RNN) and random forests. After checking grammatical rules through a syntactic parser, the meaning of those words is determined in context and this phase is called semantic analysis. Words meanings are determined by the surrounding situations and discourse analysis is used to address distinct cases after conducting sentence analysis and understanding the actual meaning of a statement, it's a very challenging task in ML and NLP. Sometimes words carry dual meanings and it's not easy to understand, NLP uses pragmatic analysis to determine the intended meaning and the actual meaning of written words, sentences and documents, etc. (text).

3. Role of Machine Learning in Natural Language Processing

The processing of natural language has several phases that a machine can comprehend. These phases involve the investigation of morphology, syntax, semantics, discourse, and pragmatics. In one way or another, machine learning contributes significantly to nearly all of these procedures. From the computer system, we received data in the form of

0s and 1s. By using ASCII code, these digits can be converted into alphabets. At the morphological level, the work is to identify the words and sentences, also known as tokenization. Then, affixes are removed by stemming. After this, the next task is to check that sentences follow the grammar rule of a language in NLP. The same words may have different meanings according to context. To remove this ambiguity, a Word Sense Disambiguation task is used. In discourse analysis, semantic analysis will not give authentic meaning. At the semantic level, the meanings of words are identified by using dictionaries (Amore) (Amore). This will be tackled by deep learning algorithms and machine learning.

3.1. Existing Studies

At first, the rule-based approach, involved the development of rules to meet language processing tasks. It was a major milestone throughout the 1980s. ELIZA, a conversational software by Weizenbaum, which used pattern matching and substitution for human-machine interactions, and SHRDLU, created by Winograd, were notable systems of that era. Both programs functioned in constrained block worlds with restricted vocabularies (Brown, 2006). During the statistical revolution that followed, ML methods like decision trees were widely used in NLP research to make probabilistic judgments based on input data. There have been numerous successful research projects concentrating on the use of natural language for developer experience. An example is Microsoft's Visual Studio development environment's recommendations "IntelliSense" feature that uses NLP techniques to provide code completion (Pullum, 2015, 1945). The second example is the "Stack Overflow Assistant" project, which utilized NLP and machine learning to answer developer queries that are based on (Zhang, et al., 2016). data Recently, there has been a trend in the use of chatbots as natural language. An example of this is GPT-3 Assistant," which uses the language model to assist developers with different topics such as code debugging and library documentation (Floridi & Chiriatti, 2020). Moreover, there are several other significant examples of natural language interfaces for developer experience, such as CodeGenie, CodeFlow, and DeepCode (Lemos, Bajracharya, Ossher, & Coatta, 2007). There has also been research on the use of natural language interfaces in other aspects of the development process, such as gathering demands and managing projects (Czerwonka, Greiler, Bird, Panjer , & Coatta, 2018).

3.2. Contribution of Studies

In recent years, Natural Language Processing and Machine Learning become more important subfields of AI, finding applications in different types of businesses. In the method called machine learning, computers can solve problems that were never meant for them to solve. Because of the many applications in real life that make this almost impossible, it is extremely difficult to explicitly program or develop an algorithm that can prepare for all possible input kinds and manage the task at hand. Artificial intelligence has two types machine learning and deep learning. Recently, it has gained attention because machine learning enables computers to learn and improve from time to time without being programmed. On the other side, deep learning is part of machine learning that uses neural networks to solve problems. Deep learning is like the structure of the human brain and can recognize patterns. They learn unstructured data like our brains (Kim & Woodland, 2000), (Han, Chen, & Zhang, 2019).

J. Kim et al. (Gui, Wang, & Zhang, 2021) show the effectiveness of handwritten dictionaries and grammar for easily deployed NLP applications, the study presents a rule-based method for evaluating intentions and entity recognition in NLP interface construction. Furthermore, the research of X. Li et al. (Pilault, Elhattami, & Pal, Conditionally adaptive multi-task learning: Improving transfer learning in nlp using fewer parameters & less data, 2015) proposes a deep learning-based method for entity recognition and intention categorization in the development of NLP interfaces. To accomplish high natural language input processing precision and efficiency, this approach can use a combination of, LSTMs RNNs, and CNNs. Pilault et al. (Pilault, Elhattami, & Pal, Conditionally adaptive multi-task learning: Improving transfer learning in nlp using fewer parameters & less data,, 2020) research focuses on transfer learning for intention categorization and entity recognition in the design of NLP interfaces, using pre-trained NLP models like BERT and GPT-2 to modify and adapt to a particular NLP application. The research not only highlights the significance of pre-trained models in NLP interface advancement but further demonstrates how transfer learning could significantly enhance the correctness and effectiveness of NLP systems. A framework for combining NLP algorithms is presented by T. Nguyen et al. (Sipio, K.R., & Nguyen, 2020) to enhance the experience while creating NLP interfaces. This framework shows various NLP algorithms improve to increase the effectiveness of NLP applications. The study offers further research highlighting the significance of NLP algorithms to enhance the experience while creating NLP interfaces. Similar to this, earlier studies have made considerable use of a range of NLP technologies in Named Entity Recognition (NER) applications. To demonstrate Spacy's great performance and efficiency, Ramachandran et al. (Ramachandran & Arutchelvan, 2021) study used it for entity extraction in the medical profession.

In their study on historical textual analysis, Johnson et al. (Johnson, et al., 2021) used NLTK, noting its extensibility and versatility as important considerations in their selection. Another well-liked option is Stanford NER, which Vychegzhanin et al. (Vychegzhanin & Kotelnikov, 2019) utilized for a benchmark study to evaluate the effectiveness of various NER methods for a wide range of languages. The research validated Stanford NER's excellent accuracy and resilience, especially when processing intricate syntactic patterns. The chatbot application by Sharma et al. (Sharma, An Analytical Study and Review of open source Chatbot framework, Rasa,', 2020) used the RASA DIET architecture recently for customer support. Their research showed that RASA DIET is specifically for daily life applications that necessitate entity extraction and intent recognition. These highlight the numerous fields and uses for which these NER technologies have been effectively used.

However, our research took comprehensive viewpoints on multiple models. There are different Entity Recognition techniques: MaxEnt Classifier with NLTK, Spacy, Conditional Random Fields with Stanford NER, and RASA DIET, and seven different algorithms for the task of intent classification: K-Nearest Neighbour (KNN), Support Vector Machine SVM Classifier, Logistic Regression, Naïve Bayes, Decision Tree, and Random Forest, along with RASA DIET. This study also explores the different levels and applications of NLP. Additionally, this research includes recent advancements directions, and advantages of NLP and ML.

4. Past, Present, and Future of NLP and ML

4.1. Applications of ML in NLP

Machine learning and natural language processing have diverse applications in industries as well as other institutes such as:

1.Chatbot and Text Summarization Tools: This application is utilized in the medical field. Recently ML and NLP analysis of live chatbots and emails has assisted scientists and researchers to fight against COVID-19 as they got data from medical helplines. This helped the physicians to help the most affected patients as the system identified the symptoms from the data. Machine learning and Natural language processing also help to summarize and categorize the patients' information to make profiles that can be easily accessed when required.

2. Semantic Analysis: FinTech uses semantic analysis where affective analysis of commercial data using NLP/ ML helps financial institutions manage the risk, check the money laundry, check the market trends, and make decisions.

3. Text Mining: The Insurers use text mining which helps to identify fraud claims and false entitlements via deep analysis of customer chats and reviews.

4. Web Scrape: Applications in supply chain areas for manufacturers. It detects the weak areas of the supply chain where more attention is required and helps the manufacturers fill the spaces where logistic improvements can take place. Manufacturers use web scrapping where online sources are scanned to make comparative data for scaling the prices and opportunities.

5. Sentimental and Social Media Analysis: Applications for benefiting the brands and companies; Companies can make their decisions by analyzing their customers' sentiments via comments and reviews. This helps them to enhance their product quality and become more powerful in the market.

6. K-Nearest Neighbor ML: ML performs a great role in video and image recognition, stock analysis, and handwriting detection. The K-Nearest Neighbor is used as a tool for this type of application (Samek).

7. Logistic regression: This model of ML/NLP helps in voting applications, churn classification, and weather classification.

8. Random Forest: Social media analysis, performance score analysis, and anomaly or fraud detection are also carried out using ML/NLP. The decision-making tool helps in the procedure.

9. Support Vector Machine (SVM): SVM is used for finance and stock market suggestions and air quality control systems.

10. Naive Bayes Theorem in ML: Text classification, disease prediction, sentiment analysis, and spam filters use the Naive Bayes Theorem in carrying out the tasks (Jiao, 2020).

4.2. Machine Learning Algorithms

Algorithms for machine learning are computer techniques created to assist machines in learning and enhancing their performance on a task through data or experience. Without being explicitly written with specific directions these algorithms employ statistical approaches to automatically identify patterns, provide forecasts, or enhance decision-making processes. Over time computers adapt and generalize data to improve their capacity to carry out tasks and make judgments. The classification of Machine learning algorithms differs based on the relationship between the independent as well as dependent variables.

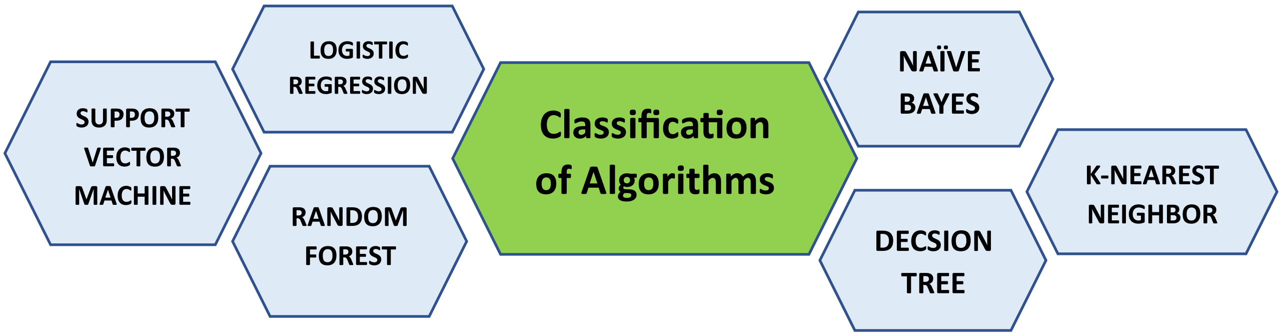

Figure 03: Classifications of Algorithms

4.3. Naive Bayes

Naive Bayes is considered a probabilistic classifier which is based on Bayes' theorem. Bayes' theorem calculates probabilistic classification using prior knowledge and the naive hypothesis where the input features show strong independence from each other. For example, with an apple as input, the classifier can classify it as fruit, again color and round shape, but they will all be attributed independently. It is the most widely used, simple, and fast algorithm of machine learning classification which works efficiently even on small amounts of training data. It is particularly suitable for the classification of text tasks like multiclass complications. This model functions under a few assumptions such as the features like words in the case of text are conditionally independent

4.4. Mathematical

$P(A|B) = P(B|A) * P(A) / P(B)$

- $P(A|B)$ represents the probability of the event (A) and event (B)
- $P(B|A)$ represents the probability of the event (B) and event (A)
- A's probability is denoted by $P(A)$.
- B's probability is denoted by $P(B)$

Real-life examples and uses of Naive Bayes are sentiment analysis, classification of text, prediction of disease, and spam filters.

4.5. K-Nearest Neighbour (KNN).

K-Nearest Neighbour classifies Models Machine learning as instance-based and works as a slow learner. A slow learner means that it never builds internal models. Instead, this classifier represents data in an N-dimensional space with n features. The main feature and advantage of the KNN classifier is its robustness against noisy data. Despite of high cost of computation, it is successfully used in various industries.

4.6. Decision Trees

It is the algorithm of machine learning used for classification models that aims to generate a tree structure to visualize the decision-making model. The hierarchy is based on equal, inclusive, and conflicting conditions. Model branches are created by dividing the data set into subsets by identifying the most important features in the classification. The tree is built using a top-down iterative divide-and-conquer method. As we move down, the training set is eventually connected to an additional decision tree. The final hierarchical structure looks like a tree with nodes and leaves. The decision tree hierarchy is easy to interpret and visualize. With little data preparation, decision trees work well for complex problems such as pattern recognition, data mining, and clinical risk identification.

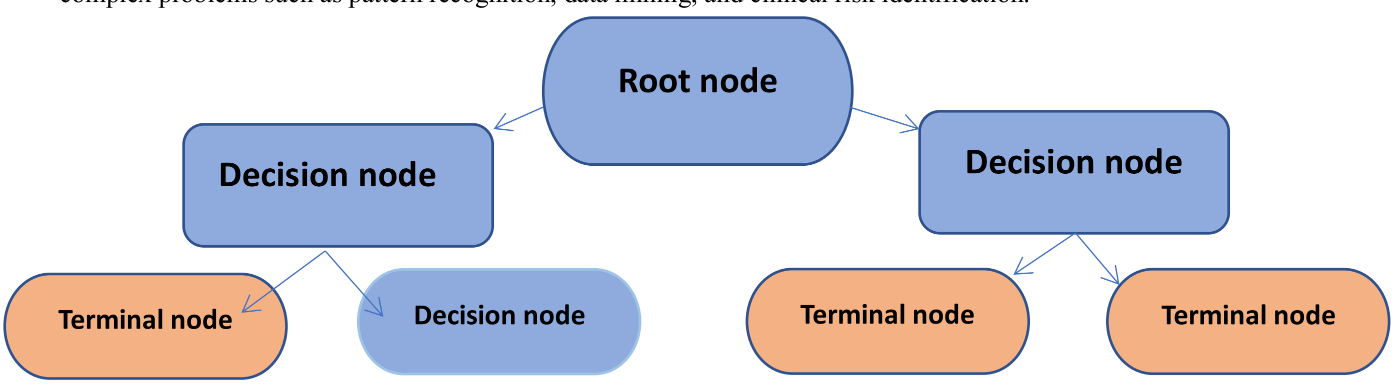

Figure 04: Root Node

4.7. Logistic Regression

Logistic regression is a model that generates class-label predictions using a sigmoid function to find the dependency between the output and input variables. It uses multiple independent variables to determine an outcome. The sigmoid function yields a probability output, which is then compared with a pre-defined threshold to assign the label to the object accordingly. It fits suitably for binary classification algorithms.

The major drawback of logistic regression is that it assumes data to be free from missing values, and the predictors are independent of each other. Depending upon the nature of the output variable, the logistic regression algorithm is categorized as Multinomial, Binary, or ordinal logistic regression.

4.8. Support Vector Machine (SVM)

The training data opinions are characterized by a support vector machine in space where the two classes are divided by a gap. The goal of the classifier is to locate in N-dimensional space a hyperplane that assigns each of the two classes to the data points. The closest data points from each class to the chosen, hyperplane can be used to define the margin. The category of data point is predicated upon the entrance of a new point. It is a non-probabilistic binary classifier. This classifier is memory efficient and very effective in high-dimensional spaces.

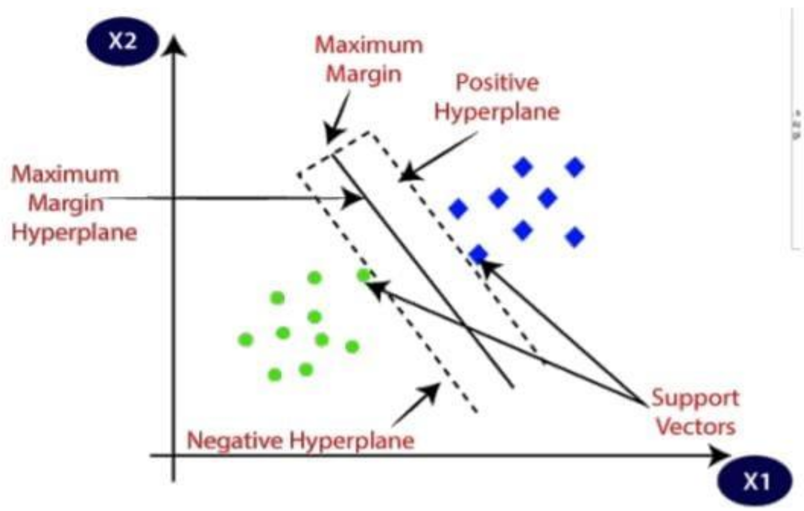

Figure: 05

4.9. Random Forest

A collaborative learning practice called random forest classification is fundamentally an ensemble of decision trees. The results of multiple forecasters are combined through ensemble learning. It is considered a descriptive estimator that usually takes into account the number of trees in subsamples of different training datasets. Hence the accuracy of predictive modelling is enhanced by using its average. During the phase of training the classifier builds several decision trees producing an output class that represents a way to predict regression for individual trees.

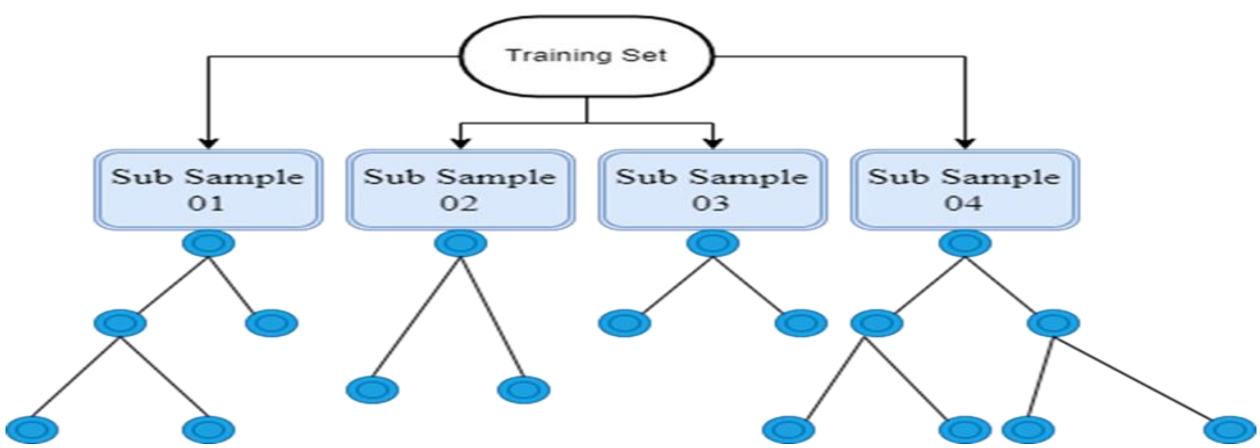

Figure 06: Training Set

5. Challenges and limitations of machine learning in natural language processing (NLP).

5.1. Data distortion

When applying machine learning to natural language processing it is one of the most important aspects then the immediate challenge is the issue of data distortion. The results that occur of machine learning models could be biased because the algorithms could learn from biased data and then continue to propagate those biases. A sentiment analysis model that was trained on biased data for example can lead to incorrect conclusions when applied to certain groups of people. This can happen because specific groups of people are more likely to express certain opinions. Another challenge is the need for these models to be explainable. Due to the complexity of the models themselves, this can be difficult to understand the decision-making process underlying NLP models. If the capacity to explain things will be preserved, it is imperative to ensure that the model is capable of making accurate and impartial judgments.

5.2. Interpretation Of the Research Findings

The study also acknowledges the limitations that shape the scope and interpretation of the research findings. Despite a complete examination of many algorithms of NLP for entity extraction and classification of intent along with an analysis of performance in executing the interfaces of natural language within low and no-code software systems numerous limitations predictably plague the investigation and will affect It is worth noting that Nokia's low-code no-code system is in under study and is in the development stage and lacks sufficient data for robust model training. Confidentiality clauses prevent the use of Nokia's proprietary data resulting in reliance on limited publicly available network data which is rare for a given industry and use case. The Lumi Nile dataset is insufficient in size and attempts to artificially increase it result in reduced data quality, leaving the final dataset smaller. The data obtained lack prior interpretation, requiring a time-consuming manual interpretation process and challenging the generalizability of the results. Another obstacle is the absence of already trained models that are suitable for a specific case which limits the application of transfer learning. Finally, a deliberate focus on simple NLP algorithms limits the exploration of the potential benefits offered by more complex deep learning models (Upreti.A, 2023).

5.3. Contextual Ambiguity

Various applications of NLP present emerging challenges including contextual ambiguity where the same words or phrases can convey diverse meanings within a single sentence. Synonyms are problematic to treat as humans use diverse words to define the same idea which raises the intricacy of language processing algorithms so working on cohesion where words are expressed in similar ways but with different lists is problematic for applications like speech-to-text and questions answering particularly when are presented in formats like no-text. The understanding of sarcasm in sentences dares NLP approaches as they can be understood differently by humans. Uncertainty in sentence understanding allows for several meanings and requires better accuracy in NLP. The foundations of informal language such as language of specific culture and idioms propose broadly valid and used models complicated and necessitating continuing training on varied sources of data. The challenges linger such as phrases or words taken with diverse meanings across different activities. The words that are misused and misspelled pose different difficulties despite the advances in the correction of grammar and self-correction of many applications that need a cautious understanding of the intentions of the author while considering the field, geographic distinctions, and linguistic intricacies. While NLP models excel in widely used languages there is still a need for universally applicable models that accommodate diverse linguistic skills and technical settings. The literature is related to discovering many challenges as well as offers a lot of appreciated comprehension. The current focuses are on the visual reasoning with NLP which usually offers auspicious but at the same time many challenging boundaries (Khurana D. K., 2023; Khurana, Koli, Khatter, & Singh, 2022).

6. Recent Advancement and Directions

In NLP and ML, we have advanced in the form of transformer models, multilingual models, and continual learning. Today we have a rapid evaluation of NLP and ML and on various applications and their transformative impact. Here we have some significant advancements in NLP and ML.

Transformer Model and NLP Advancement: The rise of transformer-based models such as GPT-3 and BERT their an impact profound on NLP tasks and the advancement in NLP and ML. These models have language improvement understanding, generation, and translation improvement.

Multilingual/ Multimodal Approach: It investigates the integration of NLP with different models such as images, and videos and it enables it to make a more comprehensive understanding of the data. It explores the applications like image captioning and it summarizes the video showcasing the synergy between NLP and computer vision. Recent approaches in NLP have models that can have the ability to translate multiple languages without any specific training and these approaches keep focus on the development of these models.

Transfer Learning and Pre-trained Modals: It examines the trend of using pre-trained modals in NLP and ML. It gives information about fine-tuning pre-trained models for performing different specific tasks that have become common practice and loading to improve efficiency and performance.
Continual Learning: It gives ongoing efforts that involve developing MT systems that can be adapted and learn from new data that ensure they stay relevant in a constantly evolving linguistic landscape.
Future Directions of NLP and ML: Recent advancements speculate on the future of NLP and ML to consider them as emerging trends like quantum computing, federated learning, and continual learning (Amini, 2023). In addition, potential application areas such as healthcare, finance, and personalized AI assistance with these recent advancements are discussed (Carrio, 2017).

7. Advantages of NLP and ML

Automate tasks: Machine learning algorithms automate language-related tasks, reducing human effort and time (Goyal, 2023). This automation impacts customer service (chatbots), data entry (speech-to-text), and content moderation (automatic flagging of inappropriate content (Sharma, An Analytical Study and Review of open source Chatbot framework, Rasa,', 2020).
Improve user experience: Better NLP enables better user experiences. Voice assistants auto-complete suggestions and personalized content recommendations enrich user interactions with technology, increasing satisfaction and engagement (Jabeen, 2023).
Extracting insights from big data: ML in NLP allows the analysis of huge amounts of textual data. Businesses gain valuable insights from customer reviews, social media sentiment, and market trends, which help in making informed decisions and formulating strategies (Zhou, 2019).
Advances in health care: NLP powered by machine learning helps analyze medical records, understand patient symptoms from text, and extract information from research papers. It helps in diagnosis, treatment planning, and drug discovery (Shatte, 2019).
Performance in education: NLP helps personalize learning experiences through adaptive learning platforms, automated grading, and intelligent tutoring systems. Meets students' individual needs and improves learning outcomes (Mosavi, 2020).
Making financial decisions: Sentiment analysis of financial news and social media data helps predict market trends and make investment decisions. Machine learning algorithms process vast amounts of text data to guide financial strategies (Amini, Mahyar, & Ali Rahmani, Achieving Financial Success by Pursuing Environmental and Social Goals: A Comprehensive Literature Review and Research Agenda for Sustainable Investment., 2023)
Legal assistance and compliance: NLP helps in analyzing legal documents, reviewing contracts, and monitoring compliance. Algorithms streamline processes, extract important information, and help ensure compliance with regulations (Amini, Mahyar, & Bozorgasl, A Game Theory Method to Cyber-Threat Information Sharing in Cloud Computing Technology, 2023).
Social and cultural influences: ML in NLP affects language use and cultural understanding. It influences language development, colloquial adoption, and cultural communication by facilitating language translation and understanding.
Ethical considerations: Machine learning algorithms raise ethical concerns about bias, fairness, and privacy. Ensuring fairness in algorithms, protecting user data, and removing biases in language models are important implications of deploying responsible AI (K. P. Johnson, 2021; Johnson, et al., 2021).
Change of job: With increased efficiency, ML in NLP also impacts job roles. It automates certain tasks, which may lead to changes in job descriptions and the need to improve skills to adapt to the changing work environment.

8. Discussion

Machine learning algorithms play an important role in natural language processing (NLP) by allowing computers to interpret and reproduce human language. These algorithms learn patterns and structures from data and provide machine translation for sentiment analysis, text summarization, etc. It allows them to perform various NLP tasks, such as the impact of these algorithms, which will profoundly revolutionize the way we interact with search engines through the development of technology. Chabot Assistants, Virtual Assistants, and Healthcare. These algorithms support vector machines (SVM), Naive Bayes, and neural networks that classify text (such as spam/spam and positive sentiments/negatives). It also applies to the models being part of the original data to make predictions on a new text. Algorithms identify and classify entities in text (such as names, dates, and locations) using techniques such as conditional random fields (CRF) or recurrent neural networks (RNNs). Statistical and neural machine translation models (such as transformer models similar to GPT) learn from parallel texts for translation between languages. Continuity-based models such as recurrent neural networks (RNNs) and transformers generate human-like text, helping chatbots summarize and answer questions. Algorithms learn word structure and probabilities in sentences, which is important for predictive text auto-completion and grammar correction. ML algorithms often use deep neural networks (DNNs) to convert speech into text by recognizing patterns and phonemes in audio data. These algorithms understand pronouns and their referents within the text, which helps preserve context. ML models, especially those based on transformer architectures, process and understand large amounts of text to generate accurate answers to questions. Extractive or abstract summarization algorithms extract the most important information and condense large texts into a condensed version. Algorithms understand the meaning behind words, sentences, or documents, which helps with tasks such as information retrieval, paraphrasing, and semantic matching.

9. Key Findings

Machine learning algorithms in natural language processing have achieved transformative results on various fronts. Through cutting-edge techniques such as feature extraction using word embeddings and transformers, these algorithms have dramatically improved the representation of words and sentences, leading to increased context understanding. Sentiment analysis capabilities enable companies to gauge public opinion, while language generation models such as GPT have demonstrated human-like text generation capabilities. ML algorithms excel at feature extraction and representation of textual data. Techniques such as word embedding (Word2Vec Glove) and contextual embedding (BERT GPT) capture semantic and syntactic information, enabling algorithms to understand the nuances of language. Deep learning models, especially neural networks, have greatly improved NLP tasks. The emergence of compiler architectures such as BERT GPT and their variants has led to breakthroughs in language understanding generation and translation.

Pre-trained language models are becoming important. By leveraging transfer learning, these models are fine-tuned for domain-specific tasks, reducing the need for large-scale labeled data and dramatically improving performance in various NLP applications. Models such as encoder-decoder architectures and the transformer-based model have revolutionized tasks such as machine translation summarization and chatbot development by processing entire sequences of text simultaneously. The introduction of attentional mechanisms into models such as transformers has increased understanding of context in language. This procedure enables the model to focus on relevant parts of the INP. Integrating multiple modalities, such as text, images, and audio, has become an emerging trend. Machine learning algorithms now facilitate tasks that involve understanding text alongside other types of data, such as image annotation or video summarization. The emergence of large language models has drawn attention to ethical considerations. Issues related to bias in training data, ethical use of language models, and potential societal impacts are actively discussed to ensure the responsible dissemination of NLP techniques. Machine learning algorithms are constantly evolving to adapt to new language differences, slang, and changes in language use. The implications of machine learning go beyond just understanding language; they affect social, ethical, and technical domains, shaping how we interact with information and the world as a whole.

10. Conclusion

The proliferation of web-enabled technologies has gained great attention. People from different geographical areas are using social networking websites for sharing opinions, sentiments, and appraisals and these contents over social networking sites have great importance to observers and analysts. The analyst is curious about knowing the positive, and negative, likes, and dislikes of their products. So, keeping in view NLP researchers machine learning models and AI systems can be utilized to make valuable decisions concerning different products and real-life entities. The comprehensive review highlights the importance of machine learning (ML) algorithms in advancing natural language processing (NLP), maintaining their significance in various domains. Through a systematic literature review, the researchers underscore the diverse applications and techniques of ML in NLP, encompassing supervised, unsupervised, semi-supervised, and reinforcement learning methods. The interconnectedness of ML algorithms with morphological, semantic, syntactic, pragmatic, and discourse analysis in NLP is elucidated, showcasing their versatility and adaptability. The study aims to explore the role of machine learning as well as Artificial intelligence-based systems to unfold the insights of machine learning algorithms. The study showcases the widespread utilization of ML/NLP tools across different sectors, including chatbots, text summarization tools, web scraping, sentiment analysis in social media, stock market analysis, disease detection in the medical field, and spam detection.

By analyzing past, present, and future patterns, this research offers insights into the progression of machine learning in natural language processing (NLP), emphasizing its contributions to technological advancements and its capacity to adjust to evolving language subtleties and modifications. Additionally, the review conducts an in-depth examination of various machine learning algorithms such as logistic regression, support vector machines (SVM), Naive Bayes, K-Nearest Neighbor, and decision trees, clarifying their roles and applications through illustrative examples. Challenges such as data distortion, interpretation of research findings, and contextual ambiguity are acknowledged, highlighting areas that require further research and development. Overall, the findings underscore the transformative potential of machine learning in NLP, paving the way for groundbreaking search engines, algorithms, and techniques. The discussion elucidates the advantages and recent progressions, emphasizing the pivotal role of machine learning in enhancing comprehension, adaptation, and innovation in the processing of virtual text.

References

Amini, M. A. (2023). "How Strategic Agility Affects the Competitive Capabilities of Private Banks". *International Journal of Basic and Applied Sciences*, 8397-8406.

Amini, Mahyar, & Ali Rahmani. (2023). Achieving Financial Success by Pursuing Environmental and Social Goals: A Comprehensive Literature Review and Research Agenda for Sustainable Investment. *World Information Technology and Engineering Journal 10.04*, 1286-1293.

Amini, Mahyar, & Bozorgasl, Z. (2023). A Game Theory Method to Cyber-Threat Information Sharing in Cloud Computing Technology. *International Journal of Computer Science and Engineering Research 11.4*, 549-560.

Amore, R. (n.d.). Speech and Language Processing Paul and Sharma, Deep Learning for NLP. *Journal of Artificial Intelligence Research (JAIR), by ACL Jurowski and Martin*.

Bansal, S. (2023, January 25). *Classification in Machine Learning: Types and Methodologies*. Retrieved December 27, 2023, from Analytixlabs: https://www.analytixlabs.co.in/blog/classification-in-machine-learning/

Brown. (2006). The encyclopedia of language &linguistics. Amsterdam Boston: Elsevier.

Carrio, A. e. (2017). "A review of deep learning methods and applications for unmanned aerial vehicles". *Journal of Sensors*.

Czerwonka, J., Greiler, C., Bird, C., Panjer , L., & Coatta, T. (2018). Code flow: Improving the code review process at Microsoft: A discussion with Jacek Czerwonka, Michaela Greiler, Christian Bird, Lucas Panzer, and Terry Coatta,'. *Vol 16. no. 5*, pp. 81–100, .

Floridi , L., & Chiriatti, M. (2020). Gpt-3: Its nature, scope, limits, and consequences,' Minds and Machines, *30*(4), 681–694.

GOYAL, A. A. (2023). *The Role of Machine Learning In Natural Language Processing and Computer*.

Goyal, A. A. (2023, May). The Role of Machine Learning In Natural Language Processing and Computer Vision. *Iconic Research and Engineering Journals, 6*(11), 185-195.

Gui, T., Wang, X., & Zhang, Q. (2021). Participation in the semeval 2015 challenge–. *in Proceedings of the 9th International Workshop on Semantic Evaluation*, 311–314.

Han, K., Chen, J., & Zhang, H. (2019). Delta: A deep learning-based language technology platform, doi:1908.01853

Jabeen, S. e. (2023). "A review on methods and applications in multimodal deep learning". *ACM Transactions on Multimedia Computing, Communications, and Applications*, 247-273.

Jiao, Z. e. (2020). "Machine learning and deep learning in chemical health and safety: a systematic review of techniques and applications". *ACS Chemical Health & Safety*, 316-334.

Johnson, K., Burns, P., Stewart, J., Cook, T., Bernier, C., & Mattingly, W. (2021). Classical Language Toolkit: An NLP Framework for Pre-Modern Languages,. *the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations, v1*, pp. pp. 20–29. . doi:10.18653

K. P. Johnson, P. J. (2021). The Classical Language Toolkit: An NLP Framework for Pre-Modern Languages. *Association for Computational Linguistics*, 20-29.

Khurana, D. K. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia tools and applications*, 3713-3744.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022, July 14). Natural Language Processing: state of the art, current trends and challenges. *Springer*, 3713-3744. Retrieved from https://link.springer.com/article/10.1007/s11042-022-13428-4

Kili. (2023). *Kili Technology Data Labeling.* Retrieved December 27, 2023, from Kili Technology: Natural language processing (NLP) is a field within computer science that works to provide computer systems with a human-like understanding of spoken language and digital text (Kili, 2023). As to dealing with human languages in spoken or text forms data

Kim, J., & Woodland, P. (2000). A rule-based named entity recognition system for speech input. *Sixth International Conference on Spoken Language Processing.*

Lemos, O., Bajracharya, S., Ossher, J., & Coatta, T. (2007). Codegenie: Using test-cases to search and reuse source code,' in Proceedings of the twenty-second IEEE/ACM. *International conference on Automated software engineering,* 525–526.

Mosavi, A. e. (2020). A comprehensive review of deep reinforcement learning methods and applications in economics.

Pilault, J., Elhattami, A., & Pal, C. (2015). Conditionally adaptive multi-task learning: Improving transfer learning in NLP using fewer parameters & fewer data. doi:2009.09139
Pilault, J., Elhattami, A., & Pal, C. (2020). Conditionally adaptive multi-task learning: Improving transfer learning in NLP using fewer parameters & fewer data, doi:2009.09139
Pullum, G. (2015, 1945). *Philosophy of linguistics,‖ The Cambridge Companion to History of Philosophy, vol*.
Ramachandran, R., & Arutchelvan, K. (2021). Named entity recognition on bio-medical literature documents using the hybrid based approach,' J Ambient Intell Humaniz Comput,., (pp. 1–10). doi:10.1007/s12652- 021-03078
Samek, W. E. (n.d.). "Explaining deep neural networks and beyond A review of methods and applications.". *Proceedings of the IEEE*, 109.
Sharma, R. (2020). An Analytical Study and Review of open source Chatbot framework. *International Journal of Engineering Research*.
Sharma, R. (2020). An Analytical Study and Review of open source Chatbot framework, Rasa,'. *International Journal of Engineering Research*, *V9.* doi:10 . 17577
Shatte, A. B. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine 49.9*, 1426-1448.
Sipio, C., K.R., K., & Nguyen, P. (2020). Democratizing the development of recommender systems using low-code platforms,' in Proceedings of the 23rd ACM/IEEE. *international conference on model driven engineering languages and systems: companion proceedings,* (pp. 1–9).
Sukhdev Singh, D. K. (2022, July 14). Natural language processing: state of art, current trends, and challenges. *Multimedia Tools and Applications*.
Upreti.A. (2023). A Comparative Analysis of NLP Algorithms for Implementing AI Conversational Assistants.
Vychegzhanin, S., & Kotelnikov, E. (2019). Comparison of Named Entity Recognition Tools Applied to News Articles., (pp. 72–77). doi:10.1109/ISPRAS47671.2019.00017
Zhang, H., Jain, A., Khandelwal, G., Kaushik, C., Ge, S., & Hu, W. (2016). Comparative Analysis of NLP Algorithms for NLI. *International Symposium on Foundations of Software Engineering*, 956–961.
Zhou, T. Z. (2019). "Big data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design. Zhou, Teng, Zhen Song, and Kai Sundmacher. "Big data creates new opportunities for materia1017-1026.