

Human Perception Based on Textual Analysis

Md. Asadul Hoque Chowdhury*

American International University-Bangladesh, Dhaka, Bangladesh

E-mail: abirchowdhury737@gmail.com

ORCID iD: <https://orcid.org/0009-0006-0312-086X>

*Corresponding author

Farhana Yeasmin Munmun

American International University-Bangladesh, Dhaka, Bangladesh

E-mail: farhanayeasminmunmun@gmail.com

ORCID iD: <https://orcid.org/0009-0000-3036-1403>

Shahidul Islam Ifte

American International University-Bangladesh, Dhaka, Bangladesh

E-mail: shahidulislamifte3@gmail.com

ORCID iD: <https://orcid.org/0009-0002-0657-7401>

Turya Gain

American International University-Bangladesh, Dhaka, Bangladesh

E-mail: joy.gain93@gmail.com

ORCID iD: <https://orcid.org/0009-0005-7000-3359>

Dip Nandi

Faculty of Science and Technology, American International University-Bangladesh, Dhaka, Bangladesh

E-mail: dip.nandi@aiub.edu

ORCID iD: <https://orcid.org/0000-0002-9019-9740>

Received: 12 May 2024; Revised: 16 July 2024; Accepted: 06 September 2024; Published: 08 December 2024

Abstract: The complex process by which humans use their senses to clarify and understand the world around them is referred to as human perception. Analyzing human perception is important for comprehension of how humans think, feel, and act, which is helpful in a variety of contexts and ultimately promotes improved understanding, communication, and engagement. This study examines the field of text mining-based human perception analysis using a precisely chosen dataset of Twitter customer service discussions. Decision Trees, KNN, Naive Bayes, and GLM are four different algorithms that are methodically examined to determine which is the most effective method for understanding and predicting human perception from textual data. After an exhaustive analysis, the Decision Tree algorithm is shown to be the best performer, closely followed by Naive Bayes. The human perception analysis of text mining, including the methodology, findings, and implications, is described in depth.

Index Terms: Text Mining, Human Perception Analysis, Algorithmic Comparison, Customer Support, Twitter Dataset, Decision Trees, KNN, Naive Bayes, GLM.

1. Introduction

In today's world of frequent improvement in technology, understanding and evaluating human behavior is vital for innovation. This establishing vulnerability concerns a variety of human activities, including mental, physical, emotional, and social. These activities, which are classified as "Human Perception" include a wide range of tasks performed by people alone or in groups, such as walking, problem-solving, expressing emotions, and opinion sharing through social networking. This work can be done by each person or in groups. Human perception analysis is an inquiry that incorporates computer science, machine learning, and behavioral sciences to research and predict human behavior based on diverse data sources. Human perception analysis is an exacerbated architecture that combines a deep learning forecaster and predictor with a Sequence-to-Sequence structure, with layers for location and attention [1]. Human perception analysis is

an exacerbated architecture that combines a deep learning forecaster and predictor with a Sequence-to-Sequence structure, with layers for location and attention [2]. The combination of cutting-edge technology emphasizes the complexity and range of analytical models in this domain. Human perception analysis aims to predict and understand the complex patterns that shape human behavior under different contexts.

Social interaction analysis alone can achieve a prediction precision of 95% for an individual's behaviors without necessitating access to personal data [3]. Activity prediction models forecast and understand how opinion dynamics impact human experiences. Furthermore, human perception is heavily impacted by the cultural and interconnected environments in which individuals inhabit. Predictive modeling and sociocultural dynamics are inextricably linked, as indicated by characteristics such as social distance, which emerge as important indications of individual as well as social attitudes [4]. Human perception prediction has applications in a variety of fields, including computer vision, artificial intelligence, machine learning, and human-computer interaction. It is critical to develop successful prediction models using behavioral patterns, historical data, and contextual clues. The cross-disciplinary integration of knowledge encourages the establishment of sophisticated models, paving the door for new applications in security systems, intelligent preferences, and healthcare. This interdisciplinary study, which is based on computational approaches and psychological principles, gives valuable insights into future behavior and has far-reaching consequences on a range of branches, including technology and relationships between individuals. Because of its many traits and complexity, human perception analysis poses challenges in successfully planning human behavior in real time. A crowdsourcing experiment revealed that text mining for human perception prediction has limits, including shortcomings in recognizing user attributes such as age and gender [5]. Privacy and ethical concerns are significant barriers, as gathering and utilizing data must respect individuals' rights.

The objective of this work is to underscore the importance of human perception analysis in understanding and predicting human perception outlining the primary challenges faced by the field, and highlighting the advancements offered by the proposed method. This multidisciplinary approach, which amalgamates deep learning with psychological and sociocultural insights, represents a significant advancement in the quest to accurately predict and comprehend human perception under various conditions and seeks to evaluate several algorithms in order to determine which is the best at accurately analyzing human perception.

The paper is divided into the following sections: The 'Introduction' studies human perception analysis based on textual representation. The 'Background' section focuses on the history and importance of understanding and analyzing human behavior in today's technological world and in the 'Method' algorithm selection, the dataset & the flow chart of the data processing are introduced. Then in 'Analysis & Discussion' the results are explained in detail with limitations, the better approach, and the consequences. Finally, computational techniques and psychological principles provide insights into future behavior and have significant implications for various industries, influencing technology and human interaction are mentioned in 'Conclusion'.

2. Background

In a time of remarkable technical progress, the capacity to anticipate human behavior is at the front of game-changing opportunities. Artificial intelligence and machine learning are revolutionizing the prediction of human behavior, offering benefits such as pattern recognition, behavior analysis, and action prediction. This predictive capacity improves effectiveness in a variety of areas and offers crucial insights to decision-makers.

2.1. Human Perception Analysis

Human perception analysis can alter problem-solving and resource allocation in an interconnected community. Predictive analytics research facilitates data-driven decision-making and lays the groundwork for future breakthroughs across a variety of industries. For instance:

The worldwide spike in crime rates emphasizes the significance of preventative measures, such as routine analysis and the identification of suspicious behaviors, which are impacted by characteristics such as security, relevance, and volitionally [6]. Human perception prediction (HPP) integrates modern technology, such as sensors and machine learning, to enhance city decision-making while also improving transportation networks, public safety, and energy efficiency, resulting in intelligent, adaptive Smart Cities. Human Perception Prediction (HPP) uses social media product assessments to improve marketing, retail, and consumer analysis, revealing feedback from consumers and stock price volatility [7]. Human Perception Prediction (HPP) is an essential technique in disaster management, assisting with evacuation planning and crowd control. It helps to understand the link between disaster risk perception and the affected population, maintaining public safety and limiting the consequences of catastrophes [8].

Human Perception Prediction enhances personal and organizational lives through predictive analytics, individualized experiences, dedicated decision-making, and optimization, promoting innovation and simplifying operations for a more intelligent, effective, and peaceful future.

2.2. Earlier Human Perception Analysis Techniques

Prior to the general adoption of modern technology, people used a mix of statistical and psychological prediction methods to predict human behavior. Understanding human behavior from a psychological perspective included decoding

social indications, language fineness, and symbolic gestures. Ekman and Friesen's 1969 study introduced the Facial Action Coding System (FACS), a systematic method for analyzing facial emotions. By recognizing and classifying certain facial muscle movements connected to various emotions,

FACS recognizes and classifies facial muscle movements associated with emotions, providing a comprehensive framework for interpreting nonverbal cues [9]. In 1955, Bales and Slater established Interaction Process Analysis (IPA), a method for evaluating interpersonal interactions in decision-making groups. The framework focuses on task-oriented behaviors, leadership, and communication. The purpose was to figure out the dynamics of group behavior and analyze interactions using patterns identified in past instances [10].

Furthermore, there is a long history of psychological investigation into written communication, such as letters and reviews. Psychology has a long history of analyzing written communication, with Pennebaker and King's 1999 study establishing new techniques for recognizing language patterns and styles, offering a sophisticated perspective on the human mind's dynamics [11].

Pioneering human behavior prediction methods have been influenced by computer science advancements like data analytics and machine learning. Advanced methods for activity prediction use data from communication channels, but understanding the historical background of manual prediction methods is crucial for comprehensive understanding.

2.3. Human Perception Analysis with Technology

Manual human perception prediction strategies have limitations like labor-intensiveness, weariness, inaccuracies, time-consuming, and high costs. They also weaken the result's credibility and introduce uncertainties due to inadequate documentation of experimental techniques [12]. Standardized experimental procedures and large-scale PR datasets are crucial for overcoming constraints in human perception prediction [13]. Technology-based methods, particularly deep learning algorithms, reduce errors and improve precision, making technology integration more efficient, dependable, and scalable for ongoing development [14].

The tech revolution has improved our understanding of human thinking through advanced technologies like text messages and social media. In 2014 research demonstrated a paradigm for long-duration complex thinking prediction, focusing on causality, context-cue, and predictability [15]. Technology-based techniques analyze complex data like wearable device sensor data, video feeds, and textual data, enabling real-time analysis for immediate forecasts and insights into human perception. These methods reduce human analysts' need for reliable forecasts, leading to industry breakthroughs.

2.4. Recent Methods of Human Perception Analysis

Human perception detection is a key area of study in computer vision and machine learning. It involves recognizing simple daily activities like walking and running, while complex ones like peeing are more challenging. Human behavior detection techniques can be categorized into unimodal and multimodal approaches. Unimodal methods depict human behaviors from a single data source, while multimodal approaches use features from multiple sources. Sensor-based approaches integrate multiple sensors for user behavior, useful in wearable technology and IoT devices [16].

The integration of wearables and the Internet of Things sensor data allows for a more precise understanding of human actions. Machine learning components like Long Short-Term Memory (LSTM) networks and recurrent neural networks are crucial for processing sequence data and human perception prediction. Deep learning, which extracts abstract features, offers outstanding performance in sensor-based activity recognition and unsupervised learning [17,18].

Computer vision and human-computer interaction techniques are used to predict human perception using noninvasive EMG data. The Optimized Noninvasive Human-Computer Interaction model accurately identifies and evaluates human movements in real-time [19]. Gupta et al. highlights the growing acceptance of vision-based approaches in HPR research, focusing on activity identification using posture estimation, wearable technology, and smartphone sensors [20]. Advances in computer science have improved the precision, accuracy, and quality of human perception prediction. Over time, advances in computer science and related technologies have enhanced the precision, accuracy, and quality of human perception prediction and it keeps on in this manner.

2.5. Human Perception Analysis Using Text Mining

Mining in computer science, rooted in machine learning and statistics, is a tool for extracting patterns and insights from large datasets. Advanced analytics, historical data, statistical modeling, and machine learning algorithms are used in data mining, web mining, and text mining to predict future events and drive innovation [21,22]. Data mining is the process of uncovering useful information buried in databases, whereas web mining is the process of locating and extracting significant information from online-related data [23]. Modern data mining focuses on computational and algorithmic challenges, focusing on scalable algorithms, pattern discovery, and text, web-based, and multimedia analysis for extracting valuable information from large volumes of textual data.[24].

Text data mining, or text analytics, utilizes advanced algorithms and linguistic analysis to analyze large amounts of text data, identifying patterns, and correlations, and facilitating informed decision-making. In 1999, Ah-Hwee Tan introduced a text-mining approach that involves refining and clarifying unstructured material to uncover patterns and knowledge [25]. Text mining has different phases of preprocessing: text data, Attribute selection, and mining patterns then comes evaluation. So, each phase is critical because correct evolution cannot occur if one stage is skipped. The

mining process might even be used to predict human perception. The imperceptibility of adversarial assaults on natural language processing models raises uncertainties concerning their relevance in real-world contexts involving individuals [26].

Text mining figures out human thought processes by analyzing unstructured data from several sources [2]. Machine-learning algorithms are utilized to extract important information, patterns, and knowledge from personal diaries, activities, and social networking sites, such as recognizing actions and detecting emotions [27]. Text mining is an excellent way to extract insights from unstructured data, but it faces problems such as context, ambiguity, and linguistic variances. The large, dynamic nature of textual data presents scalability and real-time processing issues [28]. More study is needed to focus on ethical problems such as privacy and prejudice to improve efficacy. Unstructured data, ineffective techniques, and unsupervised algorithms might cause inefficiencies in accountability, validity, or dependability [29].

Despite this, text mining's pervasive use in predicting human thinking, recognizing patterns, and extracting knowledge emphasizes its significance in data-driven the world at large. The path of text mining is dynamic, requiring continuous attempts to enhance approaches, overcome limits, and assure responsible and ethical use.

3. Method

In this study, we have followed the experimental method. Experimental text mining techniques, including sentiment analysis with additional features, are being developed to enhance accuracy and reduce implementation time in textual data [30]. Here we investigate the emerging subject of text mining-powered human perception prediction. Our study makes use of a carefully selected dataset that has been assembled to capture the complexity of human thought expressed in text. We start a thorough comparison study, using five different algorithms as our analytical instruments. With this method, we can extract the finer details and complex relationships that are weaved throughout our data, which leads to the discovery of the most effective algorithms for recognizing and forecasting human perception in the complex world of text.

3.1. Dataset

This study examines the effectiveness of text mining algorithms for human perception prediction using the 'Customer Support on Twitter' dataset. The dataset was collected from the Kaggle website [31]. The reason behind selecting this dataset is that it is a comparative dataset. In the dataset, there is a large corpus of tweets and replies from the top twenty brands that provide a focus on specific problems, fewer topics, and a wider section of natural consumers. It also offers a low message limit size for recurrent networks. There are 7 different attributes in the dataset. Those are tweet_id, author_id, inbound, created_at, text, response_tweet_id, in_response_to_tweet_id.

Table 1. The descriptions of these attributes are given below

Attribute Name	Description
tweet_id	The attribute has unique IDs for customer tweets. The anonymized IDs have been referred to as both response_tweet_id and in_response_to_tweet_id
author_id	This feature also has unique IDs for authors. This dataset has been updated with the anonymized user IDs.
inbound	This feature is important for reorganizing data for conversational model training.
created_at	This information involves the time and date the tweet was sent by the customer.
text	This feature is very important; customers are given their opinions by email in their natural language.
response_tweet_id	Customer IDs of tweets that are responses to this tweet and this is separated by a comma. This attribute has some empty values and Data is string types.
in_response_to_tweet_id	The attribute is other customer IDs which is the tweet's identifier. If there is one, this tweet is in reaction to it.

3.2. Algorithms Selection

In the field of predictive analytics, a model's ability to succeed is largely dependent on the choice of suitable algorithms. Algorithms are essential components in many applications because they operate as the computational engine that draws conclusions and patterns from data. In this part, four well-known algorithms - Decision Trees, KNN (K-Nearest Neighbors), Naive Bayes Classifier, and GLM (Generalized Linear Models), for their unique features and abilities are examined and compared. This research aims to explore the efficiency of the algorithms by evaluating the performance of these algorithms in the context of text mining-based human perception prediction. The objective is to identify the best algorithm that will enhance the accuracy of human perception analysis.

Table 2. The reasons that played a key role in choosing these algorithms are mentioned as follows

Name of Algorithm	Reason for selection
Decision Tree	Decision trees are supervised learning methods such as classification and regression that classify data based on its characteristics. This is best for Efficiency: Implementing a single decision tree speeds up processing and creates prediction rules for feature-rich datasets [32]. Structure: Creates tree-like regression or classification models using leaf nodes, branches, and root nodes representing class labels, test results, and attributes [32]. Scalability: This algorithm is faster than the Random Forest algorithm and is well-suited for handling large data sets [33].
KNN (K-Nearest Neighbors)	KNN is a non-parametric, instance-based learning algorithm that predicts based on the majority class of nearby data points. Even with huge datasets, it is computationally demanding despite being intuitive. In additionally, KNN provides Non-parametric Nature: The K-Nearest Neighbors Algorithm is a non-parametric, supervised learning classifier that uses proximity to make predictions about data point grouping [32]. Versatility and Intuitiveness: Despite its computational inefficiency and difficulty in selecting the correct K value, it offers versatility, and intuitiveness, and is a memory-based approach [32,34]. Application in Text Analytics: It is utilized in text analytics algorithms to compare K training data points and documents to determine the class of test archives [32].
Naive Bayes Classifier	Naive Bayes classifier, based on Bayes' theorem, assumes independence among predictors and performs well in complex real-world situations, particularly in text classification. It is fast, requires small training data, and efficiently handles multi-class prediction. The reason for choosing this algorithm. Efficiency in Text Classification: The Naive Bayes classifier is a supervised machine learning algorithm that is fast and efficient in predicting classes for test datasets, especially in text classification and generative learning tasks [32]. Multi-class Prediction: It is advantageous for multi-class prediction problems and performs well with less training data, assuming feature independence [32]. Probabilistic Approach: The algorithm based on the Bayesian theorem, is a probabilistic algorithm that provides excellent results in classification tasks, particularly in text data analytics [32].
GLM (Generalized Linear Models)	GLM is a versatile extension of standard linear regression that permits dependent variables with non-normal distributions. Statistical Robustness: In order to create a linear relationship between response and predictors, even when that relationship is not naturally linear, Generalized Linear Models (GLM) use a link function [32]. Overfitting Avoidance: GLM, unlike logistic regression, prevents overfitting by identifying linear relationships between input variables and supports exponential distributions and categorical predictors, making it easier for individuals to understand [32]. Model Consolidation: GLM is a statistical model that incorporates several approaches, including log-linear, ANOVA, Poisson regression, logistic regression, and linear regression, to increase its variety and effectiveness [32].

This study aims to identify the most effective method for improving predictive accuracy by analyzing these algorithms in the context of text mining-based human perception prediction. The integration of these algorithms provides a comprehensive approach to addressing the complexities in human perception analysis.

3.3. Data Processing

In order to ensure the correctness, integrity, and analytical value of the dataset, this section describes the comprehensive data processing techniques which are used. The procedure is represented in the flow chart Fig.1.

To maintain consistency for future studies, initially, we standardized the "created_at" variable's format to datetime format during the data preparation process. Then we factorized the categorical variable "inbound" into a numeric format so that we could include it in our models. To maintain data integrity, we decreased the number of rows in our dataset to one response per observation. During feature extraction, we extracted two important features: "text_length", which measures the length of customer tweets, and "response_time", which calculates the time interval between a tweet and a response to that tweet. Before model training, we split the dataset into training and testing sets using an 80/20, which allowed accurate model evaluation. This split ratio is commonly used in machine learning, ensuring enough data to train models efficiently and accurately evaluate prediction capabilities. We selected four distinct models and trained them to capture different data patterns: Decision Tree, Naive Bayes, Generalized Linear Model, and K-Nearest Neighbors. To evaluate these models, we created predictions on the test dataset and used confusion matrices to measure their accuracy and performance metrics, allowing for more informed decision-making.

4. Analysis and Discussion

Our textual analysis provided useful insights into how people share their perceptions through text, which have been covered in the part below. We used diverse techniques for evaluating model performance using visual representations like confusion matrix plots and ROC curves in addition to measures like overall accuracy (percentage of properly specified instances).

Confusion matrix plots make it possible to identify the areas where the models struggled to distinguish between different perceptual categories. These regions may highlight particular language fineness that needs more analysis. On the other hand, the model's overall discriminating capability is indicated by the ROC curves, which provide insight into

how well it can differentiate between different forms of perception. Additionally, evaluating the accuracy of the models gives a foundational comprehension of how well they extract human perception from textual input. In order to disclose the relevancy of the hidden patterns within the data, this section goes deeper into the findings.

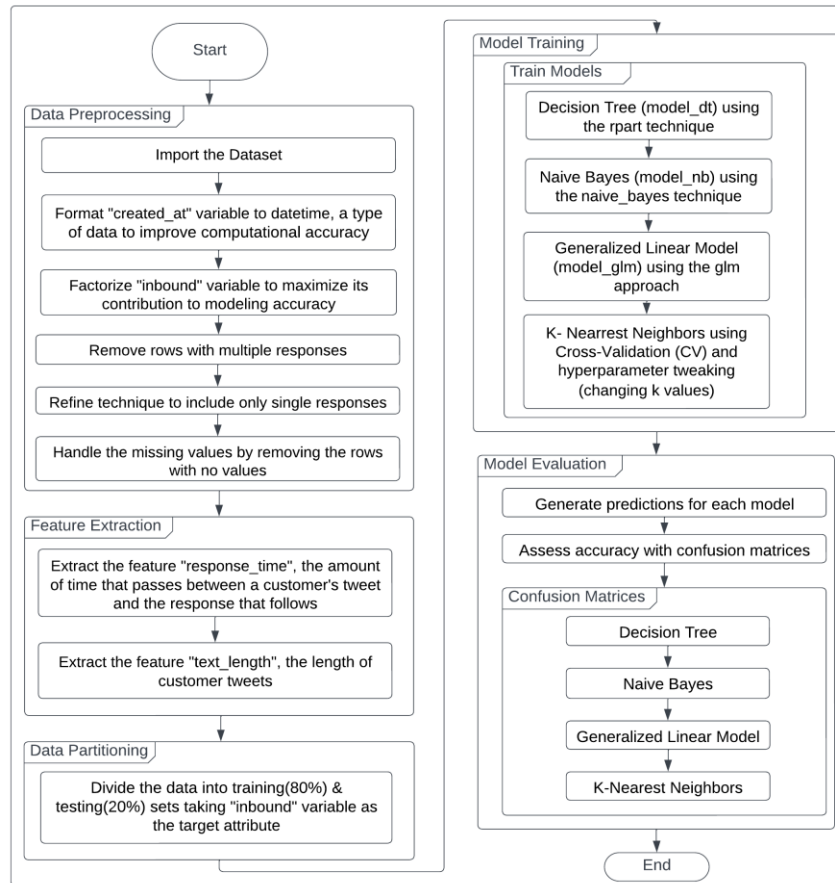


Fig.1. The data processing flow chart

4.1. Analysis

The study compares four text mining algorithms for human perception prediction on data of customer support interactions on Twitter. Decision Tree, Naive Bayes, GLM, and KNN models were compared. Each plot shows how accurate the models are in predicting events. The x-axis shows "Predicted" outcomes, which are labeled as "True" or "False," while the y-axis shows the percentage of predictions, which ranges from 0.00 to 1.00. Additionally, the color scheme uses blue for "True" predictions and red for "False" predictions.

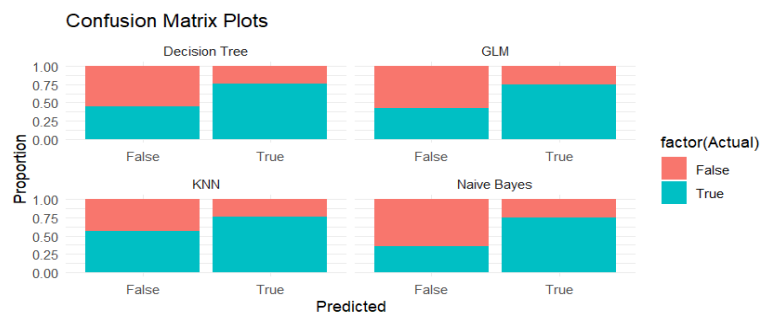


Fig.2. The confusion matrix plots for the four models

Regarding the models tested, the Decision Tree is the most successful, reflected by a greater percentage of true predictions than false ones. On the other hand, the GLM and Naive Bayes models show predictions that are equilibrium, with similar percentages of true and false results. Following that, the KNN model shows a mixed performance, with predictions that vary and show a combination of true and false classifications.

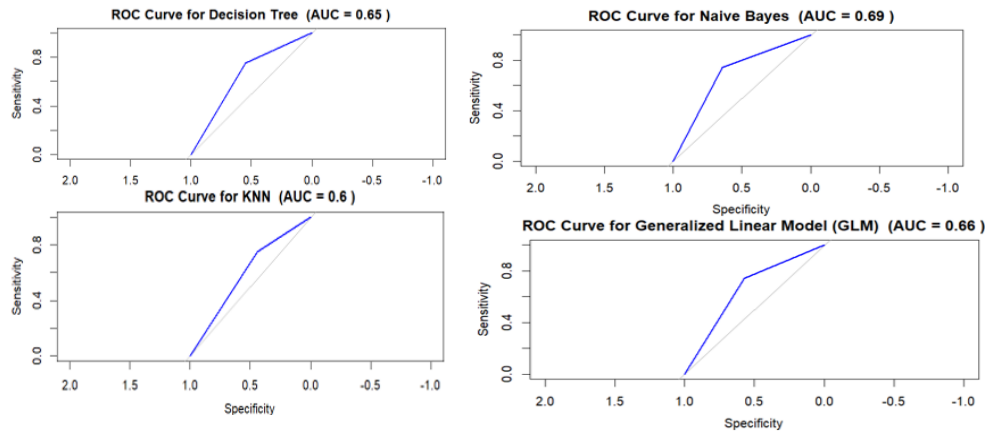


Fig.3. The ROC Curves for the four models

The ROC curve illustrates how well a binary classifier can distinguish between different classes as the threshold for classification is adjusted. The plotting is labeled as the true positive rate (Sensitivity) against the false-positive (Specificity) rate at various threshold settings. Among the presented models with an AUC of 0.69 Naive Bayes is greater. Following up GLM stands at 0.66, Decision Tree at 0.65, and KNN at 0.6. Based on these ROC curves, one would typically prefer the model with the highest AUC (Area Under the Curve) value because it represents the greatest area under the curve, reflecting a higher probability that the model will rank a random positive instance more highly than a random negative instance. In other words, it quantifies the overall performance of the classifier. Typically, an AUC closer to 1 indicates better performance, while an AUC closer to 0.5 suggests that the classifier performs no better than random guessing. Among the four models that were evaluated, Naive Bayes has the greatest AUC, which indicates that it performs the best in differentiating between the positive and negative classifications. With an AUC of 0.69, it is superior to random guessing, although it can yet be improved. The decision tree model performs somewhat worse than the Naive Bayes model, as evidenced by its lower AUC. The decision tree model is less successful than the Naive Bayes model, but it still has a decent level of predictive ability with an AUC of 0.65. The performance of the KNN model is the worst in differentiating across classes, as seen by its lowest AUC out of all the models. The KNN model's low discriminative ability is just marginally better than random guessing is shown by an AUC of 0.6. The AUC of the GLM model is greater than that of the decision tree and KNN models, but almost lower than that of the Naive Bayes model. With an AUC of 0.66, the model has modest predictive power, outperforming the decision tree by a small margin but underperforming the Naive Bayes model. However, it's important to note that the best model should be chosen based on a combination of measurements and requirements particular to a domain, not just the AUC but also its preciseness.

Table 3. The calculated accuracy for each model

Model	Accuracy (0-1)
Decision Tree	0.7442
Naive Bayes	0.7411
GLM	0.7403
KNN	0.7326

We used the 'confusionMatrix' function, which compares predicted classifications of the model against the actual classifications from the test dataset, to evaluate the accuracy of our intended models. A confusion matrix that catalogs the quantity of true positive, true negative, false positive, and false negative predictions is produced by the function. We may calculate a variety of performance measures by examining these variables, one of which is accuracy, which is the percentage of genuine findings (true positives and true negatives) among all the instances that were examined. In order to be more precise, accuracy is calculated by dividing the total number of predictions by the sum of true positive and true negative predictions. This provides us with a precise, numerical indicator of how well our model is doing at accurately identifying the test data. The decision tree's performance stands out in particular since it has a strong capacity for result prediction. Decision Tree has the highest accuracy rate of 74.42%, indicating its ability to accurately predict human perception. Naive Bayes follows closely behind with an accuracy rate of 74.11%. The GLM and Naive Bayes models, on the other hand, have a balanced prediction ability and provide reasonably accurate classifications. The KNN model is not a very strong classifier for the dataset. The KNN model has no benefit over random classifiers as it makes educated guesses about data points' labels. Naive Bayes is a robust classifier for text mining tasks, despite its simplicity and feature independence predictions.

4.2. Discussion

The GLM and KNN both algorithms showed comparable performance, but their suitability for human perception prediction may be limited. This is due to the difficulties in parameter tuning and computational efficiency. Decision Trees and Naive Bayes' (despite its simplicity and assumption of feature independence) high accuracy rates highlight the potential of text mining tools in extracting valuable insights from textual data. These findings have broader implications for natural language processing and predictive analytics, as text mining continues to evolve with advanced techniques like deep learning and sentiment analysis. Decision Trees and Naive Bayes show great promise for text mining tasks, offering high accuracy and strong performance in analyzing and predicting human behavior from textual data. Future research should aim to improve these models by integrating more detailed linguistic features and exploring ensemble methods to combine their strengths. This paper has certain limitations. One of the primary limitations was that only selective algorithms were implemented for the dataset. Due to focusing on a single dataset, the output AUC value was found to be low. Additionally, other limitations such as the study did not address areas with no replies or multiple responses.

5. Conclusions

The aim of the research is to assess the efficacy of text mining-powered human perception analysis using a real-world dataset. This study evaluates many algorithms and finds the most accurate predictor of personality characteristics and human perception. The study seeks to demonstrate the benefits of implementing modern technology and sophisticated methods in predictive modeling.

It compares four algorithms and concludes that Decision Trees are the most successful for this study. Text mining technologies produce precise and thorough forecasts of individual characteristics. Integrating cutting-edge technology and advanced methodologies into predictive models enables novel applications in a variety of sectors, including natural language processing and the investigation of unique datasets. This work significantly advances the field of text mining-based human perception prediction. It represents the potential for innovation in collecting insights and accurately estimating individual characteristics. This highlights the need to adopt cutting-edge technology and approaches when establishing models to anticipate outcomes.

The study points out that future research should use sophisticated machine learning algorithms and deep learning approaches to improve prediction accuracy, enhance the dataset to include more settings, and demonstrate practical applications in marketing, healthcare, and social media analysis. Natural language processing technologies must be refined for text mining to remain relevant and precise.

References

- [1] Jaramillo, I. E., Chola, C., Jeong, J., Oh, J., Jung, H., Lee, J., Lee, W. H., & Kim, T. (2023). Human Activity Prediction Based on Forecasted IMU Activity Signals by Sequence-to-Sequence Deep Neural Networks. *Sensors*, 23(14), 6491. <https://doi.org/10.3390/s23146491>
- [2] Gutierrez, E., Karwowski, W., Fiok, K., Davahli, M.R., Liciaga, T. and Ahram, T. (2021). Analysis of Human Behavior by Mining Textual Data: Current Research Topics and Analytical Techniques. *Symmetry*, 13(7), p.1276. doi: <https://doi.org/10.3390/sym13071276>
- [3] Bagrow, J.P., Liu, X. and Mitchell, L. (2019). Information flow reveals prediction limits in online social activity. *Nature Human Behaviour*, 3(2), pp.122–128. doi: <https://doi.org/10.1038/s41562-018-0510-5>
- [4] Farboodi, M., Jarosch, G., & Shimer, R. (2021). Internal and external effects of social distancing in a pandemic. *Journal of Economic Theory*, 196, 105293. <https://doi.org/10.1016/j.jet.2021.105293>
- [5] Flek, L., Carpenter, J., Giorgi, S., Ungar, L.H. and Preotiu-Pietro, D. (2016). Analyzing Biases in Human Perception of User Age and Gender from Text. doi: <https://doi.org/10.18653/v1/p16-1080>.
- [6] Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P. and Weller, A. (2018). Human Perceptions of Fairness in Algorithmic Decision Making. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*. doi: <https://doi.org/10.1145/3178876.3186138>
- [7] Humphreys, A. and Wang, R.J.-H. (2017). Automated Text Analysis for Consumer Research. *Journal of Consumer Research*, 44(6), pp.1274–1306. doi: <https://doi.org/10.1093/jcr/ucx104>
- [8] Schumann, J.F., Srinivasan, A.R., Kober, J., Markkula, G. and Zgonnikov, A. (2023). Using Models Based on Cognitive Theory to Predict Human Behavior in Traffic: A Case Study. [online] arXiv.org. doi: <https://doi.org/10.48550/arXiv.2305.15187>
- [9] EKMAN, P. and FRIESEN, W. 1981. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. In: Kendon, A. ed. *Nonverbal Communication, Interaction, and Gesture: Selections from SEMIOTICA*. Berlin, Boston: De Gruyter Mouton, pp. 57-106. <https://doi.org/10.1515/9783110880021.57>
- [10] Parsons, T. and Bales, R. (1955) 'Role Differentiation in Small Decision-Making Groups', in *Family, socialization and interaction process*. New York: Free Pr. u.a. Available at: <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315824307-5/role-differentiation-small-decision-making-groups-robert-bales-philip-slater>
- [11] Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6), 1296–1312. <https://doi.org/10.1037/0022-3514.77.6.1296>
- [12] Lockhart, J.W. and Weiss, G.M. (2014) 'Limitations with activity recognition methodology & data sets', *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* [Preprint]. doi:10.1145/2638728.2641306

- [13] Park, J., Song, C., Kim, M., & Kim, S. Activity Prediction Based on Deep Learning Techniques. *Applied Sciences*, 13(9), 5684. <https://doi.org/10.3390/app13095684>
- [14] Gleeson, T.A. (1957) 'ON LIMITATIONS TO PREDICTION', *Cover Journal of the Atmospheric Sciences Journal of the Atmospheric Sciences* [Preprint]. doi:[https://doi.org/10.1175/1520-0469\(1957\)014<0304:OLTP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1957)014<0304:OLTP>2.0.CO;2)
- [15] Li, K. and Fu, Y. (2014). Prediction of Human Activity by Discovering Temporal Sequence Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8), pp.1644–1657. doi: <https://doi.org/10.1109/tpami.2013.2297321>
- [16] Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). A Review of Human Activity Recognition Methods. *Frontiers in Robotics and AI*, 2, 160288. <https://doi.org/10.3389/frobt.2015.00028>
- [17] A. Zamichos, M. Tsourma, S. Papadopoulos, A. Drosou and D. Tzovaras, "User Profile-aware Daily Activity Prediction," 2023 24th International Conference on Digital Signal Processing (DSP), Rhodes (Rodos), Greece, 2023, pp. 1-5, doi: 10.1109/DSP58604.2023.10167947
- [18] Minh Dang, L., Min, K., Wang, H., Jalil Piran, M., Hee Lee, C., & Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, 108, 107561. <https://doi.org/10.1016/j.patcog.2020.107561>
- [19] Janarthanan Ramadoss, J. Venkatesh, Shubham Joshi, Piyush Kumar Shukla, Sajjad Shaukat Jamal, Majid Altuwairiqi, Basant Tiwari, "Computer Vision for Human-Computer Interaction Using Noninvasive Technology", *Scientific Programming*, vol. 2021, Article ID 3902030, 15 pages, 2021. <https://doi.org/10.1155/2021/3902030>
- [20] A. Gupta, K. Gupta, K. Gupta and K. Gupta, "A Survey on Human Activity Recognition and Classification," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 0915-0919, doi: 10.1109/ICCSP48568.2020.9182416
- [21] Surbhi Anand, Rinkle Rani, "Data Mining Types and Techniques: A Survey", *International Journal of Research in IT & Management*, Vol. 2, No. 2, 2012
- [22] Williams, G.J. (2006). *Data Mining: Theory, Methodology, Techniques, and Applications*. [online] Google Books. Springer Science & Business Media. Available at: https://books.google.com.bd/books/about/Data_Mining.html?id=dXo2g3kFxC&redir_esc=y [Accessed 18 Nov. 2023]
- [23] S., Vijayarani, Mrs., M., Muthulakshmi. (2013). Comparative Study on Classification MetaAlgorithms. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(8), 1768-1774
- [24] Smyth, P. (2000). Data mining: data analysis on a grand scale? *Statistical Methods in Medical Research*, 9(4), pp.309–327. doi: <https://doi.org/10.1177/096228020000900402>
- [25] Tan, A.H., 1999, April. Text mining: The state of the art and the challenges. In *Proceedings of the pakdd 1999 workshop on knowledge discovery from advanced databases* (Vol. 8, pp. 65-70)
- [26] Dyrnishi, S., Ghamizi, S. and Cordy, M. (2023). *How do humans perceive adversarial text? A reality check on the validity and naturalness of word-based adversarial attacks*. [online]arXiv.org. doi: <https://doi.org/10.48550/arXiv.2305.15587>.
- [27] Kumar, L. and Bhatia, P.K., 2013. Text mining: concepts, process and applications. *Journal of Global Research in Computer Science*, 4(3), pp.36-39
- [28] Atefeh, Khazaei, Ghoujdi. (2013). Exploit Prediction and Vulnerability Clustering Using Text Mining
- [29] IJSREM. (n.d.). *Using Text Mining Techniques for Extracting Information*. [online] Available at: <https://ijsrem.com/download/using-text-mining-techniques-for-extracting-information/> [Accessed 19 Nov. 2023]
- [30] Cheng, C.-H. and Chen, H.-H. (2019). Sentimental text mining based on an additional features method for text classification. *PLOS ONE*, 14(6), p.e0217591. doi: <https://doi.org/10.1371/journal.pone.0217591>.
- [31] <https://www.kaggle.com/datasets/thoughtvector/customer-support-on-twitter>
- [32] UNext. (2021). *Text Mining Algorithms: A Comprehensive Overview (2021)*. [online] Available at: <https://u-next.com/blogs/data-science/text-mining-algorithms/>
- [33] Sharma, A. (2020). *Random Forest vs Decision Tree | Which Is Right for You?* [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/#:~:text=A>
- [34] www.sciencedirect.com. (n.d.). *Text Mining Algorithm - an overview | ScienceDirect Topics*. [online] Available at: <https://www.sciencedirect.com/topics/mathematics/text-mining-algorithm>

Authors' Profiles



Md. Asadul Hoque Chowdhury is passionate Software developer about delving into the field of Artificial Intelligence and Software Engineering. His degree in Computer Science and Engineering, Bachelor of Science (BSc), is from American International University-Bangladesh (AIUB). He is dedicated to making a significant contribution to the field of Artificial Intelligence and Software Engineering through inventive research and development initiatives.



Farhana Yeasmin Munmun graduated from American International University-Bangladesh (AIUB) with a Bachelor of Science (BSc) in Computer Science and Engineering. She is particularly interested in cutting-edge technology, and she concentrates her study mostly on Data Science, Machine Learning, and Artificial Intelligence. Her passion lies in finding innovative solutions to challenges in these fields, aiming to advance technology across various industries and significantly impact the rapidly changing field.



Shahidul Islam Ifte a recent graduate of the Bachelor of Computer Science and Engineering from American International University-Bangladesh (AIUB). His major is in Software Engineering & currently wants to pursue a Master's degree. His expertise spans computer vision, image processing, machine learning, data training, agriculture, and IoT.



Turya Gain obtained his Bachelor of Science in Computer Science & Engineering from the American International University-Bangladesh (AIUB). Currently, he is doing his internship at a renowned company named Concord Enterprise. His major is Information Systems. His upcoming plan is to do a Master of Science in Cyber Security, his research interest is in the Cyber Security domain.



Dr. Dip Nandi is an associate dean in the Faculty of Science and Technology at American International University-Bangladesh (AIUB). Prior to this, he was the Director of the Faculty of Science and Technology and Head of the Undergraduate Program in the Department of Computer Science at AIUB. PROF. NANDI HAS A PhD from RMIT University in Australia, an MSC from the University of Melbourne, and a BSC from the American International University in Bangladesh. His research interests include E-learning, Software Engineering, Data Mining, and Information Systems.

How to cite this paper: Md. Asadul Hoque Chowdhury, Farhana Yeasmin Munmun, Shahidul Islam Ifte, Turya Gain, Dip Nandi, "Human Perception Based on Textual Analysis", International Journal of Intelligent Systems and Applications(IJISA), Vol.16, No.6, pp.84-93, 2024. DOI:10.5815/ijisa.2024.06.05