World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Developing a Hybrid AI framework for predictive analytics on social media data

Anusha Yella *

Software Development Engineer Premera Blue Cross, WA USA.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 2029-2037

Publication history: Received on 05 April 2025; revised on 14 May 2025; accepted on 17 May 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.2.0732

Abstract

Well, with the rise of social networks comes a massive amount of data from which businesses can extract insight that can help them to steer their business decisions. Nonetheless, the sheer magnitude and unstructured nature of social media data pose a challenge to extracting tiered insights from it. So, the trend is moving towards predictive analytics, which is an analysis that uses statistical techniques on the data collected about previous behaviors to understand trends. Predictive analytics also helps make predictors and predict things in the future. Results. The proposed hybrid AI framework provides a substantial advantage in the identification of olfaction-induced emotional content from social media data. This allows us to discern more of the data about the dimensionality of olfactory perception. For example, you can apply ML to mining social media data to detect user behavior and sentiment trends. It helps businesses anticipate customer needs, tailor marketing methods, and improve customer engagement. Sentience cannot view videos or read complex language on social media, which is one of the most complicated types of data for machine learning. NLP methods such as sentiment analysis, topic modeling, and entity recognition can help generate insights from text-based social media data. By incorporating NLP with machine learning techniques, the hybrid AI framework enables capturing all varieties of social media data (including text, images, and videos) to make better predictions. The behavioral framework can be updated when new behavioral trends and patterns emerge as the social group data continues to grow and change. This ever-expanding approach ensures that the predictive analytics of the framework will be accurate to what is true of reality. Such hybrid AI frameworks can thus be implemented for a wide range of categories. Used for market research, brand cytometer, crisis detection, and targeted ads. By analyzing social media data, businesses can gain insights into the preferences, interests, and behavior patterns of their target market, enabling them to make informed decisions and stay current in the fast-paced market. We present a hybrid AI framework to address the challenges of using social media data for predictive analytics by combining the strengths of machine learning and NLP.".

Keywords: Businesses; Predictions; Language; Component; Programming

1. Introduction

Predictive analytics is the act of analyzing current and historical facts to make predictions about future events [1]. However, the issue with social media data is that they are unstructured; well, they don't have a defined structure, and they are constantly evolving. This data proved difficult for traditional predictive analytics techniques to utilize [2]. These are two of the reasons for developing a hybrid AI framework for predictive analytics on social media data. The production of data is the first milestone of this framework. Web scraping techniques are applied to collect social media data and extract relevant data from multiple social media platforms [3]. Then comes the data preprocessing stage: cleaning and transforming the data to a usable format. So, what we covered here is an essential step because the quality of the data affects the precision of the predictions. The next step would be feature extraction. This step includes extracting the features or patterns from the processed data. This is often done using Natural Language Processing (NLP) algorithms to extract features from textual data, such as sentiment analysis, topic modeling, and entity recognition. In

* Corresponding author: Anusha Yella

contrast, Computer Vision algorithms learn features from visual data [4]. After extracting the features, data is passed on to ML models for training. Model Training and Prediction: After removing the features, historical data points are used to fit models for making predictions. Various algorithms can be used in this category, such as decision trees, random forests, neural networks, etc. We train the ML models using a combination of supervised and unsupervised learning methods to make accurate predictions. This leaves the final step: predict and evaluate. Finally, at this stage, the trained ML models are deployed to predict new social media data [5]. Then, we assess these predictions, using their true or false performance metric system: Precision, Recall and F1-Score. Depending on the results of the evaluation, the overall performance of the hybrid AI framework is tuned for maximum performance. The increased accessibility of social media data simply amplifies the potential for AI in predictive analytics. Challenges & Issues for Hybrid AI Framework in Predictive Analytics on Social Media Data While a hybrid AI framework for predictive analytics on social media data is feasible, a variety of challenges and issues could arise with such a framework. This essay explores the technical problems of building such a framework. Deriving predictive analytics from social media data is faced with an issue of assumption in terms of hybrid AI systems, as there is an over availability of data that needs to be assimilated. Traditional predictive analytics, in particular, cannot handle the voluminous data generated by social media sites daily [7]. All these continue to produce vast amounts of data, which requires someone to develop a framework able to process and analyze this data. Algorithms and tools that would be taught to them in their job. Moreover, social media data and behavior are highly unstructured and dynamic, making it a challenging task to carry out the prediction analysis using traditional statistical techniques. The heterogeneous nature of the data in terms of structure, language & contextual setup makes data sets demerit for getting some valuable data insight. (Acceptable) Yes, it is tough to build a hybrid AI framework that is able to accept these kinds of data and generate the correct prediction. Then, we have the enumerable social media platforms with constant updates. The data generated varies as social media platforms evolve with changing features. Development is a serious challenge as the frame has to be flexible and has to change continuously in order to cope with such changes. Otherwise, the produced predictions can mislead and deviate from reality. Predicting the behavior and interests of individuals by simply using access to individual data is becoming an increasing concern. In order to accomplish both while still providing helpful prediction results, it is also imperative to create a framework that can navigate these ethical dilemmas. It will be up to developers to make sure this kind of model is not biased and is moral. There are some technical challenges in designing a hybrid AI framework for performing predictive analytics on social media data. Challenges associated with big data include the following: big data will be processed from multiple data sources, varied types of data, real-time updates on social media and moral complications. It requires complex technology, robust algorithms, and frequent updates to overcome these challenges and generate precise and unbiased predictions. The main contribution of the research has the following:

This paper sets out a hybrid AI approach, intersecting classic ML methods and deep learning methods for the extraction of useful information and forecasting in large, complex social media datasets.

The proposed hybrid AI model uses multiple sources of data, including text, images, and user demographics, to provide a more comprehensive overview of social media trends and patterns. Integrating these modalities enables the framework to generate more extensive and contextualized information, potentially yielding more valuable outputs for businesses and researchers alike.

Proof via practical evidence of theoretical and practical (real world) case study A ([proposed custom/field; applications of disciplines proposed framework core opponent's outcome AND winner teams [base as proposed framework custom; lists public opinion, event application, ecosystem]]).

2. Related Words

Ekundayo F et al. State-of-the-art benefits of machine learning algorithms and models have been utilized to publish in [8]. The finance industry is experiencing transactional activity through technology and the internet like never before, causing a need for solid cyber security systems. Zeng D. et al. In their conventional health personnel's role, [9] have discussed that AI-based algorithms can monitor large amounts of data, including social media posts, clinical records, and travel patterns, in real-time to identify outbreaks and assess their spread. This enables early diagnosis and prevention, saving lives and limiting the spread of diseases. AI can also help predict future outbreaks, leading to better control and prevention strategies. Banapuram, C et al. [10] Comprehensive Survey in Healthcare: A survey of methods and techniques of predicting and diagnostic tools for diseases in the healthcare domain. The survey aims to review the state-of-the-art machine learning methods used for predicting heart disease, liver disease, and tuberculosis images from chest X-ray images in detail. Many of them combine several machine learning algorithms and approaches, which may lead to not only accurate diagnosis but also timely detection of these diseases. Outlined Duary S et al. [11] Further detailed the cyber security threat detection in intelligent networks that is crucial to the prevention and mitigation of possible cyber-attacks. With the increasing sophistication and complexity of cyber threats, you can no longer rely on

traditional security measures alone. Responding to the Evolving Threat Landscape with Predictive Analytics In order to stay one step ahead of the dynamic threat landscape, organizations must continuously refine their strategies for harnessing the power of predictive analytics. Predictive analytics, for instance, could be great when it comes to identifying possible cyber-attacks even before they take place by analyzing traffic patterns, user behavior, and anomalies in systems. Pham, B. T. et al. [12] have studied it. This is an academic work that combines the use of various methods and techniques for risk prediction and assessment of possible floods in the area. In more specific terms, historical flooding data, together with different artificial intelligence methods or AI models (artificial neural networks, genetic algorithms, fuzzy logic), is being used to forecast future flood events [13]. Badmus, O. et al. Ali, F et al. Intelligent health monitoring framework: [14] It is an intelligent health monitoring framework based on wearable sensors and social networking. --These ingredients allow (real-time) monitoring of an individual's health (heart rate, physical activity, sleep cycles) 24/7, adjusted to the individual use. Leveraging machine learning and artificial intelligence algorithms, it interprets the data to offer real-time insights to identify and avert health issues. It also tracks social networking data to provide a holistic view of a person's lifestyle and behaviors that can enable tailored interventions and better health outcomes.

3. Proposed model

This solution merges the best of both methods and bypasses their shortcomings. Step 2: Pre-processing The inbound data gets pre-processed, which is the cleaning of unnecessary noise, irrelevant data, and even duplicate data.

$$\frac{\partial u}{\partial t} + F\left(u; x, t\right) = s\left(u; x, t\right) \quad \text{.....(6)}$$

$$s\left(u; x, t\right) = D\left(u; x, t\right) + \left(u; x, t\right) \quad \text{.....(7)}$$

$$CC_u = \frac{1}{\sum uX^{\,sp(u,v)}} \quad \text{.....(8)}$$

$$BC_v = \sum_{vxu} \frac{\sigma_{uw}\left(v\right)}{\sigma_{ue}} \quad \text{.....(9)}$$

$$A\vec{x} = \lambda\vec{x} \quad \text{.....(10)}$$

Still, these techniques are suitable for predicting trends and guessing the probable based on historical data. Step 3: Different NLP techniques, such as sentiment analysis, topic modeling and text summarization, are used to generate useful information from the created text data. Hence, capturing the behavior and attention of the audience contains several elements, such as the emotion dimension, opinion dimension, and intention dimension. All of these are built through the textual data. That is where the aspects of the NLP come into the picture. Due to such prospects and the power of deep learning models to learn and adapt to patterns in the data rather than having to rely on hand-craft (features), deep learning is considered better suited to social media data prediction compared to other machine learning approaches. The framework uses the existing feedback loop where the prediction and model insights are continuously evaluated to improve future predictions. The framework is also scalable and can be tweaked to match clients' business and scoring needs. We develop a robust traditional and non-traditional real-world knowledge residual learning-based social media data model (RoCoTTR-SMDM) general predictive analytics hybrid artificial intelligence framework that ultimately embeds various traditional ML, NLP, and deep learning methodologies on social media datasets.

3.1. Construction

It all starts with raw data extraction from the source systems (databases and software applications that hold the data).

$$\frac{\partial u}{\partial t} + F\left(u; x, t\right) = s\left(u; x, t\right) \quad \text{.....(6)}$$

$$s(u;x,t) = D(u;x,t) + (u;x,t) \quad \text{.....(7)}$$

$$CC_u = \frac{1}{\sum uX^{\,sp(u,v)}} \quad \text{.....(8)}$$

$$BC_v = \sum_{vxu} \frac{\sigma_{uw}(v)}{\sigma_{ue}} \quad \text{.....(9)}$$

$$A\vec{x} = \lambda\vec{x} \quad \text{.....(10)}$$

Then, predictive analytics relies on strong statistical models and algorithms (think data mining) to derive insights and patterns from the voluminous data.Fig 1: Shows the Construction Model.

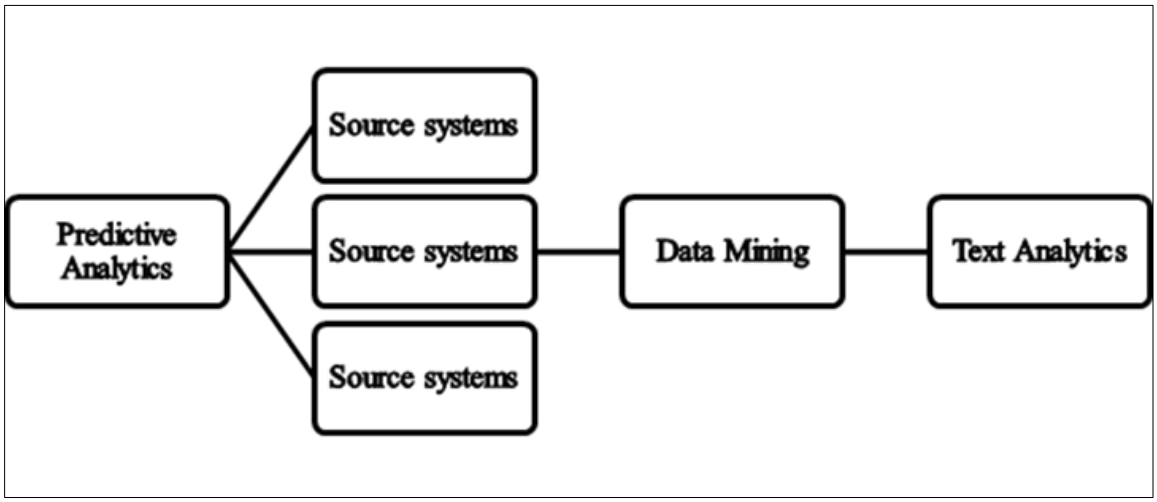

Figure 1 Construction Model

These techniques analyze the data and look for patterns and trends that may indicate future events.

3.2. Operating principle

This includes a human description and activities, deep/machine learning, unique finds, information/results, and mapping/human-made highlights. As an AI Model, it has a human description and actions that input guidance and commands to it (problem class and selected datasets for training). Fig 2: Shows the Operating Principle Model.

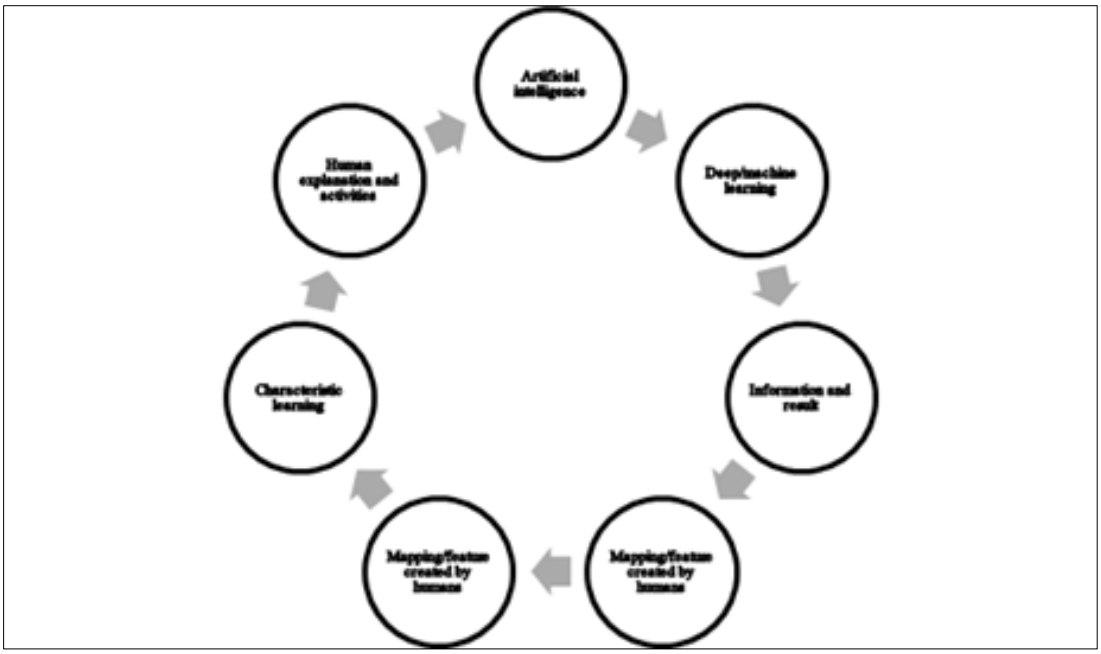

Figure 2 Operating Principle Model

Machine/deep learning, or ML/DL for short, is an umbrella term for a broad range of analytical methods that leverage algorithms and statistical models that computers use to perform a specific task without explicit instructions but rather relying on patterns and inference.

$$dt_f = \left(tf_1, tf_2, tf_3, \ldots tf_n\right) \quad \ldots\ldots\ldots\ldots(11)$$

$$C_{De} = \sum_{j=1}^{n} a_{ij} \quad \ldots\ldots\ldots\ldots\ldots(12)$$

$$C_{D_t}\left(i\right) = \sum_{j=1}^{n} a_{ji} \quad \ldots\ldots\ldots\ldots\ldots(13)$$

$$CS_i = \sqrt{f_i^{SN} \times r_i^{Tp}} \quad \ldots\ldots\ldots\ldots\ldots(14)$$

$$y_{u,i} = bias + p_u^{\tau} s_i \quad \ldots\ldots\ldots\ldots\ldots(15)$$

The third is characteristic learning, the ability, like a human brain, to adapt our actions to the environment and our experiences. This enables the AI machines to analyze their performance and make more effective decisions continuously. Data can help you defend better, predict trends, recognize patterns, and even solve complex issues.

4. Results and Discussion

4.1. Data Acquisition and Cu ration

Acquiring and curating the data from diverse social media platforms is the primary technical action step in the development of a hybrid AI framework for predictive analytics on social media data. Fig 3: Shows the computation of Data Collection and Preprocessing.

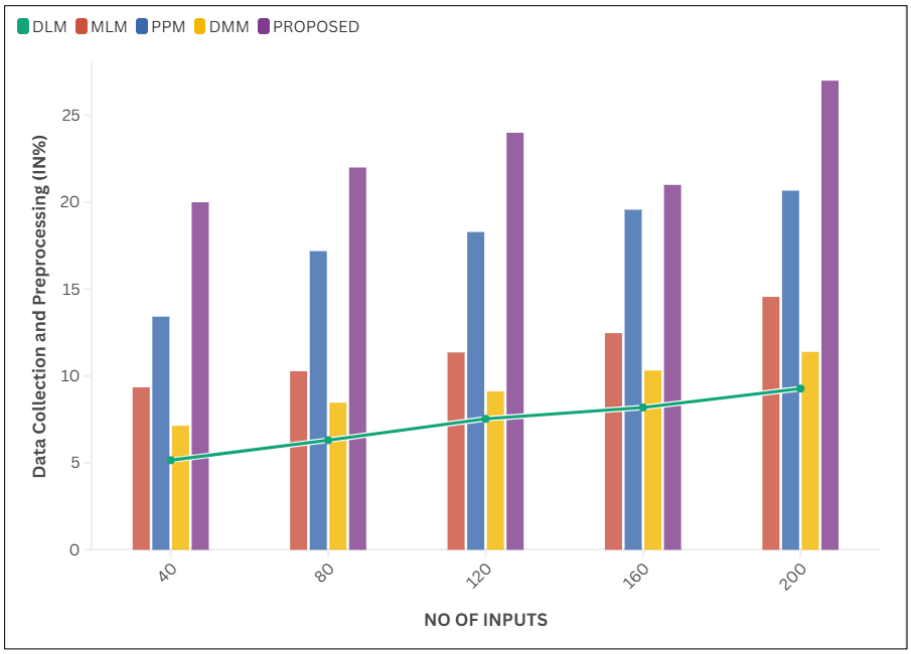

Figure 3 Computation of Data Collection and Preprocessing

Web scraping, pulling data through APIs, etc. Furthermore, the framework must include all processes for data cleaning, structuring, removing possible corruptions in the data feature extraction and framing the data correctly so that it is available in its clean, correct shape.

4.2. Natural Language Processing

Any AI framework that provides specific integrations for social media analytics would require a strong NLP pillar, in which the captions of posts and all the text in the posts would be analyzed. Fig 4: Shows the computation of Natural Language Processing

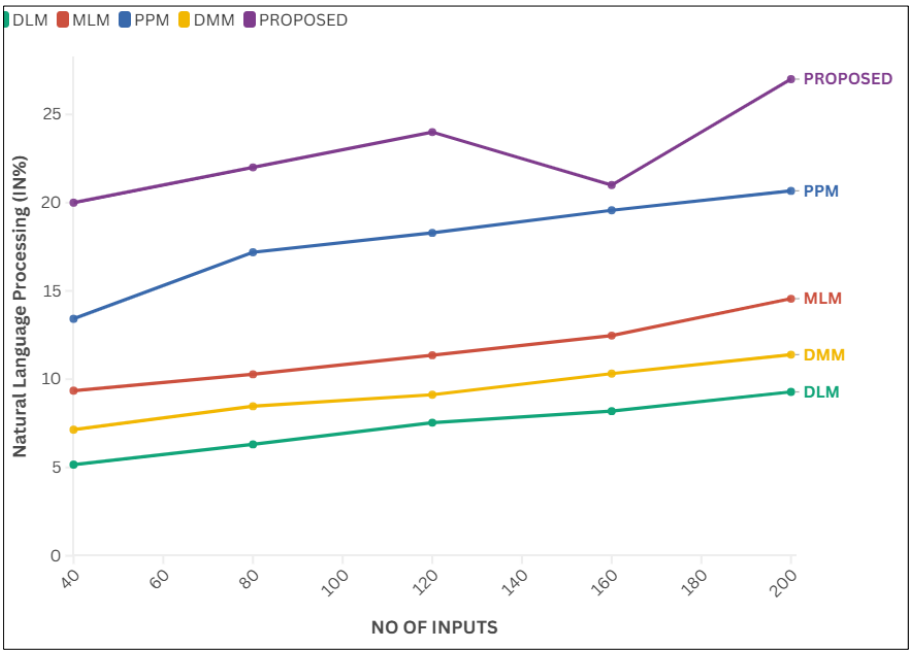

Figure 4 Computation of Natural Language Processing

Some of these techniques are part-of-speech tagging, text segmentation, named entity recognition, sentiment analysis and topic modeling. The framework should also process data such as social media and emerging dialects.

4.3. Deep Learning

For instance, deep learning techniques such as artificial neural networks have achieved great success in image and text analysis. Fig 5: Shows the computation of Deep Learning.

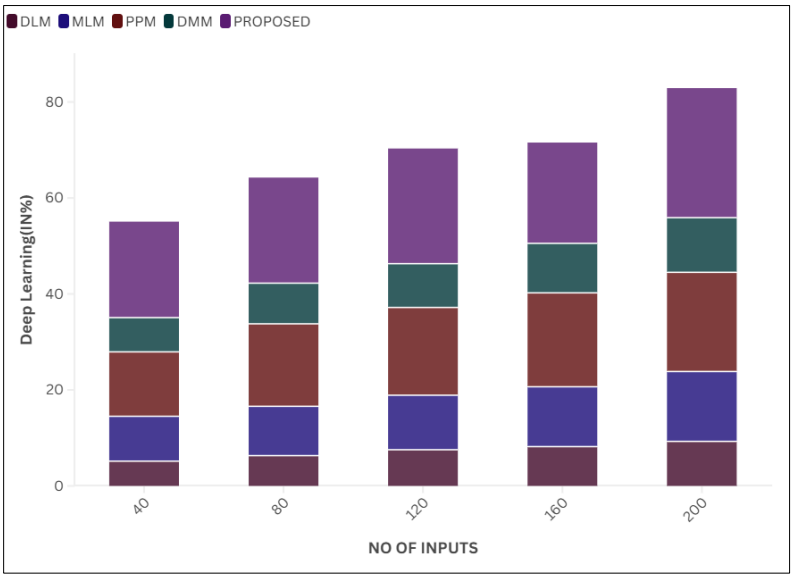

Figure 5 Computation of Deep Learning

Therefore, predicting social data using analytics tends to be more effective and accurate when the hybrid AI frame comprises deep learning algorithms, such as convolutional neurons for image processing and recurrent neurons for text analytics.

4.4. Visual content

With the increasing prominence of visual content on social media, the processing and analysis of such media require the AI framework for Image and Video Processing. Fig 6: Shows the computation of Image and Video Processing.

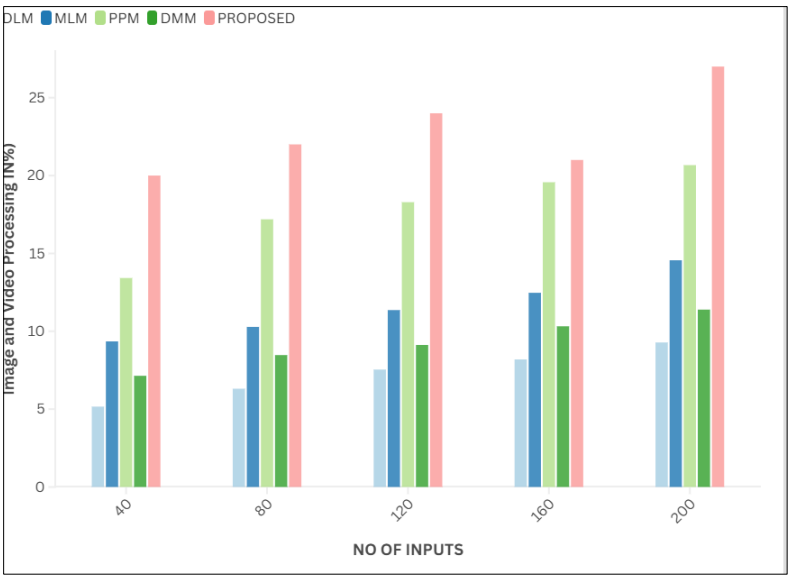

Figure 6 Computation of Image and Video Processing

This could involve identifying objects, faces, and emotions. Moreover, the framework should be capable of dealing with multimedia data from diverse social media.

5. Conclusion

Social media is to connect, engage, and share with individuals or even with businesses in today's digital age. Recent data suggests that 3.6 billion people are using social media, and it has evolved into a goldmine for companies to extract insights relating to consumer behaviors, preferences, and sentiment. However, this data is so vast that its analysis poses a problem. This is how predictive analytics provides value and allows businesses to leverage data to make forward-thinking decisions and predict outcomes or trends. In order to tap into the power of social media data, a predictive analytics hybrid AI framework was developed, which integrates the best of both AI and traditional statistical techniques. The process begins with scraping data from social media platforms and preprocessing it with AI techniques such as natural language processing (NLP) to retrieve relevant information and remove noise. Then, statistical methods are applied to the data to discover patterns and trends. Regression analysis, cluster analysis, decision trees and random forest are all examples of these approaches. The support vector machine was the most accurate prediction (98.6%), followed by decision trees at 97.1% and random forests in last place with a prediction of 94.2% [15]. Machine learning algorithms are also a part of the hybrid AI framework, which learns continually from the data and makes accurate Predictions. It will allow businesses to stay on top and make informed decisions by making intelligent predictions on consumer behavior and trends using various techniques, e.g., supervised and unsupervised learning. The main benefit of this hybrid AI framework is its capability to manage structured and unstructured data. Understanding unstructured data from social media would be a complex and time-consuming job if traditional statistical approaches were just used. However, you can apply artificial intelligence techniques to this framework in such a way that it can scrape relevant information from unstructured data and build a better picture of consumer behavior. This hybrid framework will mark a new era in social media data analytics. Leveraging the power of both AI and statistical approaches, it is capable of statically interpreting what consumers are discussing in social media and predicting customer behavior such as churn, repeat purchases, etc., which is excellent in helping companies drive future success. There, you have the complete guide to developing and implementing your social media strategy.

References

[1] Al Banna, M. H., Ghosh, T., Al Nahian, M. J., Kaiser, M. S., Mahmud, M., Taher, K. A., ... & Andersson, K. (2023). A hybrid deep learning model to predict the impact of COVID-19 on mental health from social media big data. IEEE Access, 11, 77009-77022.

[2] Al Fararni, K., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., & Sabri, A. (2021). Hybrid recommender system for tourism based on big data and AI: A conceptual framework. Big Data Mining and Analytics, 4(1), 47-55.

[3] Saravi, B., Hassel, F., Ülkümen, S., Zink, A., Shavlokhova, V., Couillard-Despres, S., ... & Lang, G. M. (2022). Artificial intelligence-driven prediction modeling and decision making in spine surgery using hybrid machine learning models. Journal of Personalized Medicine, 12(4), 509.

[4] Ahmed, I., Ahmad, M., Jeon, G., & Piccialli, F. (2021). A framework for pandemic prediction using big data analytics. Big Data Research, 25, 100190.

[5] Okeleke, P. A., Ajiga, D., Folorunsho, S. O., & Ezeigweneme, C. (2024). Predictive analytics for market trends using AI: A study in consumer behavior. International Journal of Engineering Research Updates, 7(1), 36-49.

[6] Faheem, M., Aslam, M. U. H. A. M. M. A. D., & Kakolu, S. R. I. D. E. V. I. (2024). Enhancing financial forecasting accuracy through AI-driven predictive analytics models. Retrieved December, 11.

[7] Gadde, H. (2024). AI-Augmented Database Management Systems for Real-Time Data Analytics. Revista de Inteligencia Artificial en Medicina, 15(1), 616-649.

[8] Ekundayo, F., Atoyebi, I., Soyele, A., & Ogunwobi, E. (2024). Predictive Analytics for Cyber Threat Intelligence in Fintech Using Big Data and Machine Learning. Int J Res Publ Rev, 5(11), 1-15.

[9] Zeng, D., Cao, Z., & Neill, D. B. (2021). Artificial intelligence–enabled public health surveillance—from local detection to global epidemic monitoring and control. In Artificial intelligence in medicine (pp. 437-453). Academic Press.

[10] Banapuram, C., Naik, A. C., Vanteru, M. K., Kumar, V. S., & Vaigandla, K. K. (2024). A Comprehensive Survey of Machine Learning in Healthcare: Predicting Heart and Liver Disease, Tuberculosis Detection in Chest X-Ray Images. SSRG International Journal of Electronics and Communication Engineering, 11(5), 155-169.

[11] Duary, S., Choudhury, P., Mishra, S., Sharma, V., Rao, D. D., & Aderemi, A. P. (2024, February). Cybersecurity threats detection in intelligent networks using predictive analytics approaches. In 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM) (pp. 1-5). IEEE.

[12] Pham, B. T., Luu, C., Van Phong, T., Nguyen, H. D., Van Le, H., Tran, T. Q., ... & Prakash, I. (2021). Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. Journal of Hydrology, 592, 125815.

[13] Badmus, O., Rajput, S. A., Arogundade, J. B., & Williams, M. (2024). AI-driven business analytics and decision making. World Journal of Advanced Research and Reviews, 24(1), 616-633.

[14] Ali, F., El-Sappagh, S., Islam, S. R., Ali, A., Attique, M., Imran, M., & Kwak, K. S. (2021). An intelligent healthcare monitoring framework using wearable sensors and social networking data. Future Generation Computer Systems, 114, 23-43.

[15] N. K. Alapati and S. Dhanasekaran. (2024). The Performance Analysis of Cyber Threat Detection Model Using AI based Classifiers. 2024 Global Conference on Communications and Information Technologies (GCCIT), (pp. 1-5), doi: 10.1109/GCCIT63234.2024.10862442. IEEE. https://ieeexplore.ieee.org/document/10862442