International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Novel Approach to Predicting Personality Behaviour from Social Media Data Using Deep Learning

Harshita Samota[1], Dr. Saumya Sharma[2], Dr. Humera Khan[3], Dr. M. Malathy[4], Dr. Gurwinder Singh[5], Dr. Surjeet[6], Dr. R. Rambabu[7]

Submitted: 03/12/2023 **Revised**: 26/01/2024 **Accepted**: 01/02/2024

Abstract: In the era of prolific social media engagement, understanding and predicting personal life behaviour play a pivotal role in tailoring user experiences and providing targeted services. This paper introduces a cutting-edge approach leveraging deep learning techniques to predict personal life behaviour from social media data. The proposed methodology goes beyond traditional analyses by harnessing the power of deep learning, specifically recurrent neural networks (RNNs), to discern intricate patterns within users' online activities. By training on vast datasets of opinions, sentiments, and personal activities shared on social platforms, the model establishes a nuanced understanding of individual behaviour. The study addresses the inherent challenges of capturing the dynamic nature of personal life behaviour and explores the potential of recurrent neural networks in forecasting future behaviours. To validate the efficacy of the proposed approach, comprehensive evaluations are conducted using real-world social media datasets. The results not only demonstrate the model's ability to predict personal life behaviour accurately but also shed light on the interpretability and generalizability of the deep learning framework in this context. This research contributes to advancing the frontier of predictive analytics in social media, offering valuable insights for personalized user interactions and targeted services.

Keyword: *Social Media, Deep Learning, Predictive Analytics, Personal Life Behaviour*

1. Introduction

In the ever-evolving landscape of social media, where individuals willingly share an abundance of personal information, the ability to comprehend and predict personal life behaviour has become a paramount pursuit. As users engage with platforms, expressing opinions, sentiments, and detailing personal activities, an intricate tapestry of behavioural data is woven. This phenomenon has prompted a paradigm shift in the way we approach user understanding and engagement. Against this backdrop, this research endeavors to introduce a groundbreaking approach to predicting personal life Behaviour from social media data, employing the formidable capabilities of deep learning. Social media platforms have become ubiquitous channels for self-expression, communication, and information dissemination. Users willingly divulge details about their daily lives, preferences, and sentiments, creating an unprecedented wealth of data. However, extracting meaningful insights from this vast and dynamic repository poses a formidable challenge. Traditional methods fall short in capturing the nuanced patterns and intricate interconnections that define personal life Behaviour in the digital realm.

Enter deep learning, a revolutionary paradigm in artificial intelligence that has demonstrated unparalleled success in processing complex and unstructured data. This research pioneers the application of deep learning techniques, specifically recurrent neural networks (RNNs), to discern patterns within the multifaceted landscape of social media. The objective is clear: to go beyond conventional analyses and unveil a novel approach capable of predicting future personal life Behaviour with unprecedented accuracy. At the heart of this research lies the recognition that personal life Behaviour is a dynamic and evolving entity, shaped by a multitude of factors. By harnessing the power of deep learning, the proposed model seeks to not only decipher current Behaviour but also anticipate future actions based on past engagements. The utilization of RNNs facilitates the capture of temporal dependencies inherent in social media data, allowing the

[1]*Research scholar, Banasthali vidyapeeth, Tonk Rajasthan, India*
Email: harshitasamota@gmail.com
[2]*Assistant Professor, The NorthCap University, Gurugram*
Email: Saumyasharma@ncuindia.edu
[3]*Assistant Professor, Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Kingdom of Saudi Arabia*
Email: Hkhan@nbu.edu.sa
[4]*Professor, Department of Computer Science and Engineering, Vel Tech High Tech DrRangarajan DrSakunthala Engineering College, Avadi Chennai-62*
Email: drmalathyanandan@gmail.com
[5]*Associate Professor, Department of AIT-CSE, Chandigarh University, Punjab, India.*
Email: singh1001maths@gmail.com
[6]*Associate Professor, Affiliation- Bharti Vidyapeeth's College of Engineering, New Delhi*
Email- surjeet.balhara@bharatividyapeeth.edu
[7]*Professor & HOD, Department of Computer Science & Engineering, Rajamahendri Institute of Engineering & Technology, Rajamahendravaram.*
Email: rambabureddy.rampatruni@gmail.com

model to grasp the sequential nature of user interactions and unveil latent patterns that elude traditional methodologies. The significance of predicting personal life Behaviour extends far beyond the realm of academic curiosity. The ability to anticipate user actions and preferences holds profound implications for personalized user experiences, targeted content delivery, and the optimization of services. For instance, a platform equipped with the foresight to predict a user's upcoming life events could tailor content, recommendations, and advertisements with unparalleled precision, fostering a more meaningful and engaging user experience. Throughout the course of this research, rigorous evaluations will be conducted using real-world social media datasets to validate the efficacy of the proposed deep learning model. The results are expected not only to demonstrate the model's predictive prowess but also to shed light on the interpretability and generalizability of the deep learning framework in the context of social media analytics. In essence, this research marks a significant stride toward unlocking the latent potential embedded within the vast sea of social media data. By introducing a novel approach to predicting personal life Behaviour, it aims to redefine the landscape of user understanding, ushering in a new era of tailored and anticipatory digital interactions. As we delve into the intricacies of this innovative methodology, the promise of a more insightful and responsive digital future beckons.

2. Literature Survey

1. In this paper, the authors [1] proposed a deep learning model designed to classify user interests on social networks. Employing a novel approach, the model demonstrated its effectiveness in accurately categorizing user interests highlighted the significance of leveraging deep learning techniques for enhancing the classification of user preferences in the dynamic context of social networks. The authors introduced innovative methodologies, contributing valuable insights to the field of machine learning and soft computing. The findings underscore the potential of deep learning models in deciphering and understanding user interests within the intricate framework of social network data.

2. In this, [2] the machine learning techniques employed in the study demonstrated a robust capability to forecast mental health status based on relevant data inputs. The findings of the research underscored the potential of machine learning applications in the domain of mental health, offering a new perspective on predictive models for well-being. The authors contributed to the ongoing discourse in the field by presenting a valuable framework for leveraging machine learning in mental health forecasting, potentially opening avenues for more effective interventions and support systems.

3. In this the author [3] presented a comprehensive analysis of various approaches and their effectiveness in predicting personality based on online Behaviour. The findings revealed insights into the correlation between digital activities and personality characteristics, contributing valuable knowledge to the intersection of psychology and digital communication. The author's work provided a nuanced understanding of the potential for utilizing digital footprints for personality prediction, enriching the discourse in information technology and its implications for profiling individuals through their online interactions.

4. In this study, the authors [4] published in Soft Computing, the research delved into the analysis of Twitter data to forecast election outcomes. The deep learning model showcased its efficacy in leveraging patterns and trends from Twitter activity to make accurate predictions regarding election results. The study demonstrated the potential of utilizing social media as a valuable source for gauging public sentiment and opinions related to political events. The authors' work presented a significant advancement in the application of deep learning techniques in the realm of electoral predictions, shedding light on the intersection of technology, social media, and political forecasting.

5. In this the research [5] presented an innovative approach to visual sentiment analysis. The model, incorporating event concepts and object detection, demonstrated heightened adaptability and accuracy in predicting sentiments associated with visual content in social media. The study contributed to the advancement of sentiment analysis by integrating event concepts, providing a nuanced understanding of contextual influences on visual sentiment. The authors' work offered valuable insights into the intersection of computer science, social media, and sentiment prediction, showcasing the potential for enhanced models in deciphering emotional responses to visual content.

6. In this paper, the authors [6] presented the study demonstrated the effectiveness of machine learning paradigms in capturing and analyzing patterns within community interactions to make accurate personality predictions. The authors' approach provided a distinctive perspective on leveraging social structures to enhance the precision of personality prediction models. The findings contribute to the growing field of personality prediction by showcasing the relevance of societal community dynamics in understanding and forecasting individual personality traits through advanced machine learning techniques.

7. In this paper, the authors [7] integrating social honeypots, which mimic genuine social network accounts, with machine learning algorithms, the study

demonstrated heightened accuracy in detecting and classifying social spammers. The findings highlighted the effectiveness of the proposed hybrid methodology in mitigating the impact of social spam, thereby contributing to the ongoing efforts to enhance the security and integrity of online social networks. The authors' work presented a valuable synergy of techniques to address the evolving challenges posed by deceptive activities in online social platforms.

8. In this study, the author [8] developed by the authors. Published in Computation, the research proposed an integrated methodology that blends parameter estimation techniques with machine learning algorithms. The study demonstrated the effectiveness of this hybrid approach in accurately. By combining the strengths of parameter estimation and machine learning, the authors provided a comprehensive framework for enhancing the precision of predictions related to air quality. The findings contribute to the field of environmental science and computational modeling, offering a valuable tool for assessing and managing the impact of harmful substances on atmospheric conditions. The authors' work underscores the potential of integrating diverse techniques for robust predictions in environmental monitoring.

9. In this the author proposed [9] strategy, incorporating spatiotemporal Behaviour aspects, demonstrated its efficacy in capturing the dynamics of user engagement. The study contributed to the advancement of computational social systems by presenting a refined random walk strategy that considers both spatial and temporal dimensions. The findings highlight the relevance of integrating spatiotemporal Behaviour for more accurate modeling of user interactions on online social platforms, offering valuable insights for researchers and practitioners in the field of computational social systems. The authors' work showcased a nuanced perspective on enhancing the realism of algorithms governing user Behaviour in online social environments.

10. In this research, the study introduced [10] an innovative methodology to enhance the autonomous identification of cyberbullying instances. The Transformer network-based word embeddings approach demonstrated its effectiveness in capturing nuanced linguistic patterns associated with cyberbullying, thereby improving detection accuracy. The research contributes to the field of intelligent unmanned systems by leveraging advanced natural language processing techniques to address the critical issue of cyberbullying. The findings underscore the potential of Transformer networks in developing robust algorithms for autonomously identifying and mitigating online harassment. The authors' work presents a significant step forward in deploying intelligent systems for the protection of online users from harmful Behaviours in digital spaces.

11. In this study, the authors [11] demonstrated the feasibility of predicting both home and vacation locations through the analysis of online photo collections. By leveraging advanced algorithms, the authors showcased the potential of extracting valuable insights into users' lifestyles from their digital footprints. The findings contribute to the broader field of web and social media analytics, offering a novel perspective on utilizing visual data to infer aspects of individuals' daily lives and travel Behaviours. The authors' work provides a valuable foundation for understanding lifestyle patterns through the lens of online photo sharing.

12. In this [12] social media and machine learning were utilized by the authors. The application of machine learning algorithms, particularly in conjunction with PERS, demonstrated effectiveness in accurately recognizing and categorizing both personality and emotion based on social media data. The findings contribute to the evolving field of neural computing and applications, showcasing the potential of combining social media analysis and machine learning for nuanced insights into individuals' emotional states and personality characteristics. The authors' work represents a significant advancement in leveraging technology for psychosocial analysis through the lens of online communication platforms.

13. In this paper, the authors [13] explored the concept of personal sensing, aiming to understand mental health through the integration, the research delved into the potential of utilizing pervasive sensors and advanced machine learning techniques for gaining insights into individuals' mental health. The study highlighted the significance of continuous, unobtrusive data collection through ubiquitous sensors in monitoring and analyzing various aspects of mental well-being. By employing machine learning algorithms, the authors demonstrated the capacity to extract meaningful patterns and associations from the collected data, offering a promising avenue for personalized mental health assessment. The findings contribute to the evolving field of clinical psychology by showcasing the potential of leveraging technology to enhance the understanding and monitoring of mental health through continuous and unintrusive means.

14. In this study, the author [14] identified the research focused on leveraging deep learning techniques to discern personality characteristics from users' emoji usage patterns. The study demonstrated the efficacy of this methodology in extracting meaningful insights into individuals' personalities based on their emoji preferences. By employing advanced algorithms, the

author illustrated the potential for decoding psychological traits through the analysis of emoji usage on messaging platforms. The findings contribute to the field of multimedia tools and applications by offering a novel perspective on utilizing emojis as a rich source of information for inferring personality traits in the digital realm. The author's work represents an innovative exploration of the intersection between communication patterns and deep learning for psychological profiling.

15. In this research, the authors [15] introduced an innovative methodology for enhancing POI recommendations by integrating deep learning and social network information. The hybrid model, DeePOF, showcased its efficacy in leveraging both deep CNNs and friendship connections to improve the accuracy of location-based recommendations. The findings contribute to the field of concurrency and computation, highlighting the potential of combining deep learning techniques with social network features for more refined and context-aware POI suggestions. The authors' work represents a significant step forward in developing advanced recommendation systems that consider both content features and social relationships in the context of location-based services.

ANN in Predicting Personality Behaviour from Social Media Data

In recent years, the exponential growth of social media platforms has provided an unprecedented opportunity to delve into the intricate relationship between online Behaviour and individual personality traits. Leveraging the vast amount of data generated on these platforms, researchers have turned to advanced computational techniques, with Artificial Neural Networks (ANN) emerging as a powerful tool for predicting and understanding personality Behaviours. ANNs, inspired by the human brain's neural network architecture, are computational models composed of interconnected nodes, or neurons, organized in layers.

In the context of predicting personality Behaviour from social media data, ANNs offer several advantages. They excel at capturing non-linear and intricate patterns that may be inherent in the multidimensional nature of personality traits. Social media data, often characterized by its unstructured and diverse content, presents a challenge for traditional statistical methods. ANNs, with their ability to adapt and learn from data, become invaluable in deciphering the nuanced relationships between online Behaviour and personality characteristics.

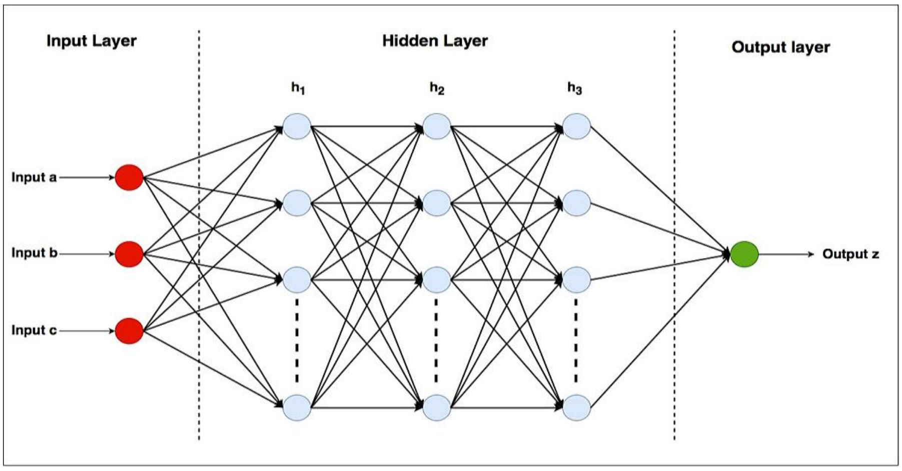

Fig1: Basic ANN Model

- The feature-rich nature of social media data, encompassing textual, visual, and temporal aspects, aligns well with the capabilities of ANNs to handle complex input structures. Natural Language Processing (NLP) techniques can be integrated into the ANN architecture to process and analyze textual content, while image recognition algorithms can be employed for visual data. The temporal dynamics of online interactions can be considered by incorporating time-series data into the ANN design. However, it is crucial to acknowledge the challenges associated with using ANNs in this context.
- The interpretability of neural network models remains a concern, and efforts must be made to elucidate the decision-making processes of the model to ensure the reliability of the predictions. Additionally, ethical considerations, such as user privacy and bias in training data, need to be carefully addressed to mitigate potential risks associated with deploying predictive models based on social media data.

3. Methodology

There is a correlation between a person's personality traits and the content of their Facebook profile. Some prediction models factor in users' and their networks' Facebook activity. Attributes that were studied were the number of likes, groups, tags, and friendship networks. All of these earlier studies demonstrate a strong correlation between demographic information and a user's Facebook profile in terms of their usage habit. In light of this, the predictive model for this study is based on a mixture of two separate predictive models from prior work and use a distinct dataset. The results of three previous research formed the basis of the model adopted. For this investigation, we selected factors that have a strong chance of influencing personality prediction.

When developing an algorithmic model, one must adhere to certain procedures. Data pre-processing is the first and most important stage. The data is prepared for the work to be carried out in this stage. Incorrectly handled data fed into the model might drastically degrade its performance or perhaps totally derail it. Data processing is required prior to classification or input into the ANN network. Afterwards, the data is transformed and eventually sent into the ANN for categorization as part of the procedure itself. The model utilised in this study is depicted in Figure 2, which is a flow diagram. This study's dataset was drawn from the myPersonality project's database, which includes Facebook information from over 4 million people who have been assigned a personality classification according to the Big Five model.

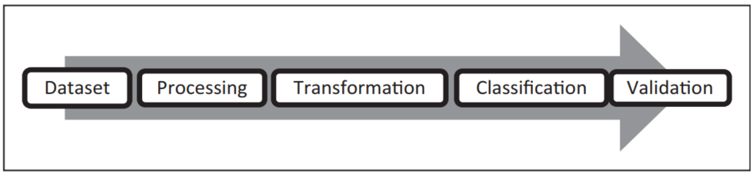

Fig 2: The classification process model

While respecting users' right to privacy, this Facebook app harvests data from their profiles and provides them with the option to complete psychometric tests that determine things like life happiness and Big Five personality traits. The datasets were constructed using the processed and analysed data that was obtained from the application. User demographics, activity, and friendship network size are all part of the data set. The datasets listed below were downloaded for the purpose of this investigation. The datasets were extracted from the myPersonality database and required merging into a single file. To do this, Microsoft SQL was employed, with each database identified by its unique user ID. The script that was used to combine the several databases is located in the Appendix.

When merging the datasets, we only kept participants who were common to all three and removed those who were missing from one or more of the others since the participation counts in each database were different. There were 1,330,543 rows of distinct participants with distinct user IDs remaining in the dataset after the data merger. The consideration of missing values was undertaken following the successful merging of several database files. Due to the potential for bias, effects on findings and generalizability as well as a significant loss of knowledge, missing data can substantially impact study outcomes. Using IBM SPSS, we performed missing value analysis to make sure the data was complete and accurate. We found that the combined file had many missing values, therefore we would need to fix them before we can utilise the database for the research.

Missing values were handled by writing another script in MSSQL. The script iterated through the columns until the percentage of missing values dropped below 10%, at which point the data was reduced to 7,443 participants, beginning with the column with the highest percentage of missing values. If a missing value is found, the row is deleted. Once the percentage of missing values was lowered to 10% or below, the remaining missing values were filled in using SPSS's replace missing value option using series mean.

It is now possible to analyse the data for inclusion in the neural network model as the dataset only contains individuals with complete data (See Figure 2). There were 4,425 females and 3,013 males out of the total of 7,439. The largest age grouping of participants was those between the ages of 18 and 25, followed by those between 26 and 40, 7.4 and less than 18, 6.4% between 40 and 60, and 0.7% more than 60. The dataset also includes the participants' Big 5 personality traits; out of the total, 96% were open, 4% were introverted, 57% were neurotic, 43% were agreeable, 9% were not, 87% were constantly aware, 13% were not, and 88% were extroverted.

Table 1: Personality distribution

Value	OPE	NEU	AGR	CON	EXT
Yes	7,181	4,209	6,789	6,502	6,567
No	257	3,229	649	936	871

Tensors containing floating point data should be used when feeding input into the neural network. Because of the potential impact on training, the data should also not span extremely broad ranges. In order for the neural network to function correctly with the dataset, it was necessary to rescale the values to make them consistent in scale. Python was also used for data rescaling because it is going to be used to build the neural network. Thanks to Google's TensorFlow, data processing and neural networks have become more easier in Python. All operations were executed with TensorFlow as the backend.

The following provided instructions for rescaling:

1. Some columns, including relationship status and gender, have to be rescaled in order to normalise the data, as they include nominal values. The input variable was extended to 18 thanks to "One hot encoder" that divided the column into 10 and scaled it into binary.

2. The rest of the data was normalised in Python using the MinMax Scale (0-1) function. Because it facilitates feedforward and reverse propagation in gradient descent calculations, it was selected.

3. Any value above 0.5 was considered a 1, while any number below 0.5 was considered a 0. The most effective way to avoid this is to convert the data into vectors ranging from -1 to 1 or 0 to 1, which will normalise the data

\Scaling some of the data from 1 to 10 is also possible. It starts with the first integer and assigns it a value of 1 while assigning the remaining integers 0s; if there are 10 variables, it repeats this process for each integer that remains. The next step is to send the normalised and transformed dataset on to classification, which will yield a value of 0 or 1 (Figure 2).

Classification has historically made use of a variety of approaches; however, this study made use of a multi-label back propagation neural network methodology. Multi-label categorization has seen extensive use of ANN. To guarantee proper generalisation and effective learning, it is necessary to efficiently pick the setup and parameters in an ANN model. The model utilises supervised theory to improve itself and updates itself using a feed-forward process using the back-propagation update approach. Although it is expensive in terms of the amount of data needed to train, this approach is multi-purpose, highly effective, and yields excellent results. Layers one through three of the output structure make up the hidden layer. Through the synapses, we collect data from all of the input neurons and then multiply it by a random weight matrix. After adding up all of the weights, the inputs are sent to the output layer via an activation function. The research employed both the ReLU and sigmoid activation functions. As a starting point, the ReLU activation function is employed.

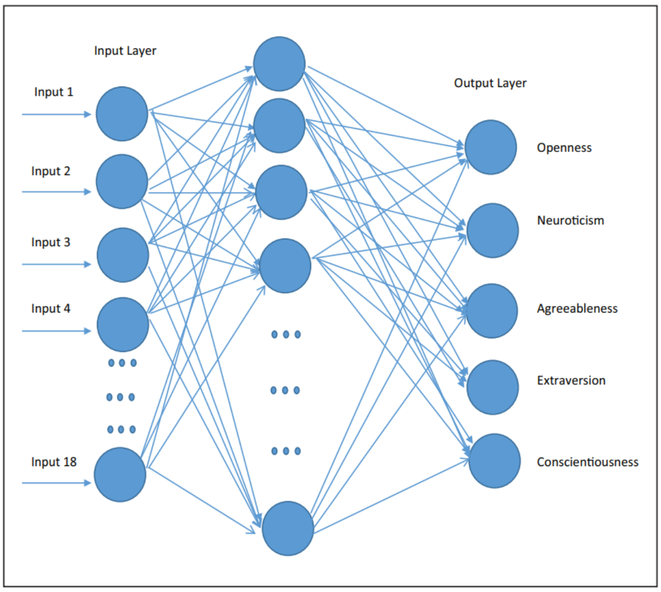

Fig 2: Neural Network Model

A multi-label classification issue is an ANN job where each label sample has the potential to have more than one label, rather than being mutually exclusive. Keras, an ANN framework, does a good job of tackling this challenge to consider, this study employs n samples (X x x = … { } 1 n,) and n labels (y y y = … { } 1 n,), where y_i ∈ {1,2,3,4,5}, and the prediction probability is P(c_j— x_i). The following step is to construct a basic artificial neural network Although it's not hard to build the input and hidden layers, the output layer and its type are crucial considerations when working with multi-label data. By default, when faced with a multi-classification issue, the softmax layer is selected. In order to break down the multi-label classification problem. Each label will undergo its own set of improvements, and these improvements will have no effect on the likelihood of any other labels. As seen in Figure 2, there are a total of 18 inputs and 5 outputs.

The research used the Google-developed Keras API, which is based on TensorFlow, to construct the ANN model. In order to facilitate better learning, this API makes it easy to build models and manipulate data. Developers may simply experiment with different optimisation strategies and algorithms because to TensorFlow's flexibility, which helps in simplifying implementation. We used Python to build this model. Because of its easy-to-read syntax and general accessibility, Python has quickly risen to the ranks of the most popular data science languages To determine and monitor the model's accuracy, classification algorithms divide the dataset into training, testing, and validation data. About 60% of the data will be training data, which serves as examples for the network to learn patterns and variations from. It will be given to the training model frequently to help it make judgements. During the training process, an additional 20% of the data is utilised to check the training quality.

4. Result

Because of its great data visualisation capabilities and plenty of machine learning tools, Python was chosen to do the data processing and modelling. The amount of hidden layers and neurons must be determined before the neural network can be constructed. It is prudent to begin with a few hidden layers, ideally one completely linked hidden layer, depending on the data (Chollet, 2017). A completely linked hidden layer was used to connect the input and output layers in the model. A non-saturating activation function called ReLU was employed between the first and hidden layers, with the sigmoid activation function serving as the final output activation function. Finding the optimal amount of hidden neurons for a given task may be challenging without first trying out a few different models. Overfitting or underfitting may occur if there are not enough buried neurons or if there are too many.

When building a neural network, finding the optimal parameters often requires a lot of trial and error. Choosing the optimal amount of hidden neurons follows a certain pattern (Heaton, 2008).

- Its size shouldn't exceed twice that of the input layer.
- Two-thirds of the total number of neurons (input and output)
- This study found the neurons with the optimum performance by trying different amounts of hidden neurons and subsequently increasing them.

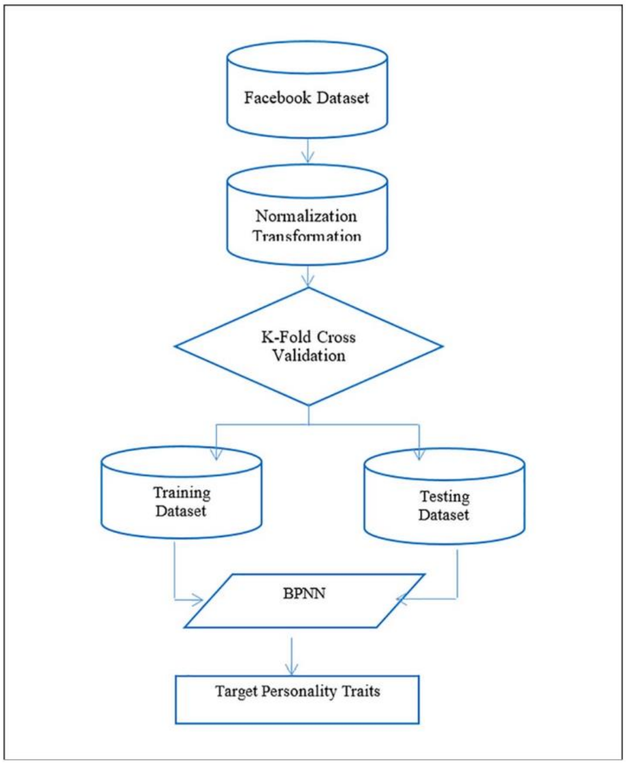

Fig 3: Work Flowchart

To begin, the data was imported using Panda's Data Frame, which offers a plethora of useful features for working with data. The following stage, following the successful import of data, is to generate a features and target variable matrix. This allows the network to recognise the input and output files, as well as the nine input variables and five output variables. As mentioned before, attributes like relationship status were classified from 1 to 10. However, keeping it this way would imply that 10 is greater than 1, which is obviously not the truth. So, it would be more accurate to say that single is higher than divorced. A fake variable called OneHotEncoder was constructed in Python using the SciKitLearn library function (Pedregosa et al., 2011) to deal with them. The SciKit-Learn package was used for transformation and normalisation after a single hot encoding. In order for the network to do a better job of classifying the data, the dataset is vectorized from 0 to 1. Additionally, the output target dataset underwent transformation with a threshold of 0.5. Another usage of SciKit-Learn is to divide the dataset into two parts: the training set and the test set.

Assessment and instruction come next. There were two distinct approaches to this problem. One method required partitioning the dataset by hand, while the other used K-Fold cross validation. The methods used in the study are shown in Figure 3.

Training

The first scheme was manually partitioned as follows: 75% for training and 25% for testing in the first part, and 67% for training and 33% for testing new data that the model has never seen before in the second portion. However, the second strategy used K-fold cross-validation, which was further divided into K-10 Cross validation and K-5 Cross validation.

The 18 characteristics were accommodated in the end phase of the BPNN model by making the input neuron consist of 18 neurons. Openness, agreeableness, neuroticism, extraversion, and conscientiousness were the five personality classifications represented by the five neurons that made up the output layer. In an effort to find the sweet spot, we tweaked a number of parameters—including learning rate, hidden neurons, and splits—and compared the outcomes, finding that they varied. There was a directive to terminate training after 10 epochs if the network's loss performance continued to decline, even though the maximum number of epochs was set to 1,000. A Windows 10 PC with an Intel Core i5, 8 GB of RAM, and a 3.30 GHz processor was used for all the computations and training. The various neural networks' training parameters are displayed in Table A1.

Testing

Both on the test data and in terms of generalizability, the model works admirably. In other words, when fed data from new Facebook Activity Users, the model produces impressive generalisation capabilities. The results of the hamming loss and prediction accuracy for all the networks that were trained. One of the measures utilised for evaluation in this study is the hamming loss. By establishing a balance between the predicted and target data, it is possible to calculate accuracy. The percentage of wrongly predicted labels is known as the hamming loss. With an optimal hamming loss of 14.96%, the model is able to accurately categorise individuals based on their Facebook activity for over 85% of the time. All of the trained networks' hamming losses and prediction accuracy outcomes from testing schemes 1 and 2 are shown in Table 1

5. Conclusion

This research set out to test how well artificial neural networks (ANNs) could categorise and forecast users' big five personality traits using information culled from their Facebook profiles. In order to properly deduce an individual's personality. The prediction accuracy reached 86.07%. Results from training and testing with fresh data demonstrated that ANN with well tuned parameters may achieve good accuracy on complicated multi-label tasks like personality categorization. The need for internet tools that can better comprehend consumers' personalities has grown in tandem with the increasing need among various firms to better understand their clientele. However, additional data may be supplied to the model to enhance its training phase. An individual's social media accounts are a great source is that it did not use additional ways to verify accuracy, including partial least squares or other machine learning techniques. To enhance prediction and allow for better comparison of outcomes, comparable research should be conducted on several accounts belonging to the same users. Lastly, further research into the use of neural networks for personality prediction and understanding is needed in order to find solutions to improve people's lives. Users will no longer have to fill out lengthy personal forms to find out their personality type; instead, Facebook users' activities can be used to determine their personality type, greatly improving the model's prediction accuracy. Users have the option to post their results publicly on their wall. A corporation's ability to anticipate its employees' personalities could help it provide better service, depending on the needs of the company.

Marketers can narrow their focus more precisely; for instance, when promoting new items, they can reach individuals with an openness personality type; while promoting security products, they can reach persons with a neurotic personality type. Additional research may be conducted by integrating artificial neural networks (ANN) with the Big Five personality traits to examine Facebook data for the purpose of predicting depressive symptoms and thoughts of suicide. The data that can be extracted from Facebook is quite rich. The next step is to gather additional data and cross-validate it with data from other social media sites outside Facebook in order to make the model more accurate.

References

[1] Dinh, X. T., & Van Pham, H. (2020, January). A proposal of deep learning model for classifying user interests on social networks. In *Proceedings of the 4th international conference on machine learning and soft computing* (pp. 10-14).

[2] Rayan, A., & Alanazi, S. (2023). A novel approach to forecasting the mental well-being using machine learning. *Alexandria Engineering Journal*, *84*, 175-183.

[3] Valanarasu, R. (2021). Comparative analysis for personality prediction by digital footprints in social media. *Journal of Information Technology and Digital World*, *3*(2), 77-91.

[4] Ali, H., Farman, H., Yar, H., Khan, Z., Habib, S., & Ammar, A. (2022). Deep learning-based election results prediction using Twitter activity. *Soft Computing*, *26*(16), 7535-7543.

[5] Fouad, Y., Osman, A. M., Hassan, S. A., El-Bakry, H. M., & Elshewey, A. M. (2023). Adaptive Visual Sentiment Prediction Model Based on Event Concepts and Object Detection Techniques in Social Media. *International Journal of Advanced Computer Science and Applications*, *14*(7).

[6] Subbarayudu, Y., Komma, B., Reddy, D. B., Raju, V. A., Ashok, K., & Shiva, B. (2023, September). An efficient novel approach for personality prediction using machine learning paradigms based on societal community. In *AIP Conference Proceedings* (Vol. 2754, No. 1). AIP Publishing.

[7] Lee, K., Caverlee, J., & Webb, S. (2010, July). Uncovering social spammers: social honeypots+ machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 435-442).

[8] Madiyarov, M., Temirbekov, N., Alimbekova, N., Malgazhdarov, Y., & Yergaliyev, Y. (2023). A Combined Approach for Predicting the Distribution of Harmful Substances in the Atmosphere Based on Parameter Estimation and Machine Learning Algorithms. *Computation*, *11*(12), 249.

[9] Fu, C., Xia, Y., Yue, X., Yu, S., Min, Y., Zhang, Q., & Leng, Y. (2021). A novel spatiotemporal Behaviour-enabled random walk strategy on online social platforms. *IEEE Transactions on Computational Social Systems*, *9*(3), 807-817.

[10] Pericherla, S., & Ilavarasan, E. (2021). Transformer network-based word embeddings approach for autonomous cyberbullying detection. *International Journal of Intelligent Unmanned Systems*.

[11] Zheng, D., Hu, T., You, Q., Kautz, H., & Luo, J. (2015). Towards lifestyle understanding: Predicting home and vacation locations from user's online photo collections. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 9, No. 1, pp. 553-560).

[12] Talaat, F. M., El-Gendy, E. M., Saafan, M. M., & Gamel, S. A. (2023). Utilizing social media and machine learning for personality and emotion recognition using PERS. *Neural Computing and Applications*, *35*(33), 23927-23941.

[13] Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology*, *13*, 23-47.

[14] Saeidi, S. (2023). Identifying personality traits of WhatsApp users based on frequently used emojis using deep learning. *Multimedia Tools and Applications*, 1-14.

[15] Safavi, S., & Jalali, M. (2022). DeePOF: A hybrid approach of deep convolutional neural network and friendship to Point-of-Interest (POI) recommendation system in location-based social networks. *Concurrency and Computation: Practice and Experience*, *34*(15), e6981.