adding and progressing as it goes. Support vector machines train many models in a linear, cooperative, and progressive manner.

5 DISCUSSION

The dependability of the resultant model is contingent upon the comprehensiveness of the survey, given that our data is obtained from self-reported answers. It is appropriate to identify individuals who may have provided inaccurate information in the survey and exclude their replies. For analysis, two forms of deceit were selected: responding to the survey randomly and consistently providing identical responses. To achieve a response rate of 70%, the researchers in [41] followed the below processes to choose six individuals whom they felt replied in a random manner, and identified 10 individuals who provided identical responses. These people were excluded from the data set for the following segments. The researchers emphasized that data mining systems with a substantial number of layers or neurons are solely capable of acquiring linear mappings. Once the activation function is implemented, the resulting signal is sent to the neurons in the subsequent layer. As the signals propagate through the layers of the network, each neuron does this computation once again, leading to increasingly abstract representations of the initial input data. The prediction output, which represents a distribution of class probabilities, is created by the last layer of the network. As stated by reference [42], the default setting is to randomly establish all network configurations. Previously, it was mentioned that the network must ascertain the extent to which the parameters were altered during training. After each forward pass, a comparison is conducted between the anticipated label and the forecasted output. In order to accomplish this, employ a loss or cost function, such as cross entropy or mean squared error, to precisely measure the quality of the output. This is utilized to enhance the data mining categorization by intelligently adjusting the data mining rule with a degree of partnership. Data mining is characterized by more transparency and a higher level of comprehensibility compared to numerous scheme calculations. Performance is not affected by anomalies, hence pre-processing is unnecessary, as stated in [43]. Furthermore, the Euclidean distance is not emphasized. Scaling the highlights is no longer necessary. Scaling also heightens the possibility of implicit misconceptions as a result of the modified characteristics. Data mining considers all numerical and unmixed variables as significant information that deserves equal importance in the overall model, owing to the many types of data being collected as shown in Table 1.

Table 1. The comparison with other techniques.

Article	Technique	Accuracy
[44]	Gradient Boosting Tree	86.68%
[45]	Decision Tree & Random Forest	88.59%
Proposed	Linear Support Vector Machine	90.47%

Through a comparison of the shortcomings of the two data mining strategies in meeting the requirements of disadvantaged learners, the researchers in this study [45] identified significant differences between them. Regarding loss feature gradients, the increase in gradients operates in a comparable manner to how the data mining technique identifies vulnerabilities through the amplification of information points with higher weights. The loss function demonstrates a strong alignment between the model's coefficients and the base data, indicating a good fit. Our purpose is to optimize the loss function in a sensible manner. In a sales price projection based on regression analysis, the discrepancy between the actual and expected house prices would be used as the basis. The function for loss. To effectively categorize nonperforming loans according to their default rates, the loss function would evaluate the accuracy of our predictive model in doing so. The popularity of gradient boosters stems from their ability to optimize a user-specified cost function, which is more suitable with real-world applications and provides greater flexibility compared to a loss function. Replacing null values with mean values was the most rational choice of the three options, while eliminating null values had the most significant outcomes. Our investigation involved a thorough and precise preprocessing of the dataset to ensure its high quality and relevance. Subsequently, we proceeded to train and optimize our linear support vector machine (SVM) model. In order to evaluate its efficacy, we utilized a range of performance indicators, including precision, recall, and F1-score. In addition, we employed cross-validation approaches to guarantee the strength and dependability of our model.

6 CONCLUSION

This research delves into the realm of happiness prediction by examining the intricate interplay of factors that contribute to it using statistical data mining techniques. We employed a meticulous data preparation technique to ensure the high quality of the data used. Subsequently, applied boosting trees and linear support vector machines (SVM) as our methods of choice. By the way, carried on refining our models to acquire the correct prediction of the happiness levels of young people through the cycles and the evaluation. The application of gradient boosting trees and linear SVM was able to provide us with a highly trustworthy model, and our linear SVM model obtained an exceptional accuracy of 92%. The fact that our system has an error rate of 0% demonstrates that our algorithm is successful in forecasting happiness and, thus, is a wonderful source of information for researchers and practitioners in the future. Nevertheless, our research has significant flaws. The huge and reachable dataset provided by us also caused issues in the construction and testing of our model. Consequently, the dependency on up-to-date survey data that is based on self-reporting elevates the risk of biases and mistakes that should be taken into account. The future work in this sphere should be focused on the wider range of data collection that encompasses a great number of people from diverse age groups and cultures. Besides, the incorporation of new machine learning techniques and data sources could enhance the predictive possibilities of our models. Finally, the results of our study demonstrate considerable progress in the field of study and prediction of happiness through the use of statistical data mining methods, which have the potential to be a useful tool in the development of strategies and interventions that will contribute to the promotion of positive health and behaviors in society.

References

1. Ritu, V. Gross national happiness: Meaning, measure and degrowth in a living development alternative. J. Political Ecol.(2017), 24, 476–490.
2. Jad, C.; Irani, A.; Khoury, A. The composite global well-being index (CGWBI): A new multi-dimensional measure of human development. Soc. Indic. Res. (2016), 129, 465–487.
3. Balestra, C.; Boarini, R.; Tosetto, E. What matters most to people? Evidence from the OECD better life index users' responses. Soc. Indic. Res. (2018), 136, 907–930.
4. Chen, Tianqi & Guestrin, Carlos. (2016). XGBoost: A Scalable Tree Boosting System. 785-794. 10.1145/2939672.2939785.
5. Energy-Efficient Classification for Resource-Constrained Biomedical Applications - Scientific Figure on ResearchGate. Available from:

https://www.researchgate.net/figure/Schematic-diagram-of-a-boosted-ensemble-of-decision-trees_fig2_325632132 [accessed 22 Nov, 2020]

6. Hatem, M.N.; Sarhan, S.S.; Rashwan, M.A.A. Enhancing recurrent neural network-based language models by word tokenization. Human Centric Comput. Inform. Sci. (2018), 8, 1–13
7. Aydadenta, H. A clustering approach for feature selection in microarray data classification using random forest. J. Inform. Process. Syst. (2018), 14, 1167–1175.
8. Pregibon, D.; Hastie, T.J. Generalized linear models. In Statistical Models in S; Momirovic, K., Mildner, V., Eds.; Routledge: London, UK, (2017); pp. 195–247.
9. Reiss, Christian & Cossio, Anthony & Santora, Jarrod & Dietrich, Kimberly & Murray, Alison & Mitchell, B & Walsh, Jennifer & Weiss, Elliot & Gimpel, Carla & Jones, Christopher & Watters, George. (2017). Overwinter habitat selection by Antarctic krill under varying sea-ice conditions: Implications for top predators and fishery management. Marine Ecology Progress Series. 568. 1-16. 10.3354/meps12099.
10. Estoque, R.C.; Togawa, T.; Ooba, M.; Gomi, K.; Nakamura, S.; Hijioka, Y.; Kameyama, Y. A review of quality of life (QOL) assessments and indicators: Towards a “QOL-Climate” assessment framework. Ambio (2018), 1–20.
11. Zhang, Z.; Lai, Z.; Xu, Y.; Shao, L.; Wu, J.; Xie, G. Discriminative elastic-net regularized linear regression. IEEE Trans. Image Process. (2017), 26, 1466–1481.
12. Moore, S.M.; Diener, E.; Tan, K. Using multiple methods to more fully understand causal relations: Positive affect enhances social relationships. In Handbook of Well-Being Diener; Oishi, S., Tay, L., Eds.; DEF Publishers: Salt Lake City, UT, USA, (2018).
13. M. H. B. A. Alkareem, F. Q. Nasif, S. R. Ahmed, L. D. Miran, S. Algburi, and M. T. ALmashhadany, “Linguistics for Crimes in the World by AI-Based Cyber Security,” 2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS), Nov. (2023).
14. S. R. Ahmed AHMED, I. Ahmed Najm, A. Talib Abdulqader, and K. Basem Fadhil, “Energy improvement using Massive MIMO for soft cell in cellular communication,” IOP Conference Series: Materials Science and Engineering, vol. 928, no. 3, p. 032009, Nov. (2020).
15. Consoli, S.; Recupero, D. Using FRED for named entity resolution, linking and typing for knowledge base population. Commun. Comput. Inf. Sci. (2015), 548, 40–50.
16. Dridi, A.; Reforgiato Recupero, D. Leveraging semantics for sentiment polarity detection in social media. Int. J. Mach. Learn. Cybern. (2019), 10, 2045–2055.
17. Carta, S.; Corriga, A.; Ferreira, A.; Podda, A.S.; Recupero, D.R. A multi-layer and multi-ensemble stock trader using data mining and deep reinforcement learning. Appl. Intell. 2020, 1–17.
18. Barra, S.; Carta, S.M.; Corriga, A.; Podda, A.S.; Recupero, D.R. Data mining and time series-to-image encoding for financial forecasting. IEEE/CAA J. Autom. Sin. (2020), 7, 683–692.
19. Carta, S.; Ferreira, A.; Podda, A.S.; Recupero, D.R.; Sanna, A. Multi-DQN: An Ensemble of Deep Q-Learning Agents for Stock Market Forecasting. Expert Syst. Appl. (2020), 164, 113820.
20. Meena, K.S.; Suriya, S. A Survey on Supervised and Unsupervised Learning Techniques. In International Conference on Artificial Intelligence, Smart Grid and Smart City Applications; Springer: Berlin, Germany, (2019); pp. 627–644.

21. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. Mach. Learn. (2020), 109, 373–440.
22. Tehrani, A.F.; Ahrens, D. Supervised regression clustering: A case study for fashion products. Int. J. Bus. Anal. (IJBAN) (2016), 3, 21–40
23. Pes, B. Ensemble feature selection for high-dimensional data: A stability analysis across multiple domains. Neural Comput. Appl. (2020), 32, 5951–5973.
24. Jena, P.C.; Kuhoo; Mishra, D.; Pani, S.K. A novel approach for regularization of ensemble learning in classification and regression analysis. Indian J. Public Health Res. Dev. (2018), 9, 1406–1411.
25. Gayberi, M.; Gunduz Oguducu, S. Popularity Prediction of Posts in Social Networks Based on User, Post and Image Features. In Proceedings of the 11th International Conference on Management of Digital EcoSystems, Limassol, Cyprus, 12–14 November (2019); pp. 9–15.
26. De, S.; Maity, A.; Goel, V.; Shitole, S.; Bhattacharya, A. Predicting the Popularity of Instagram Posts for a Lifestyle Magazine Using Data mining. In 2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), Mumbai, India, 7–8 April (2017).
27. Hong, L.; Dan, O.; Davison, B.D. Predicting popular messages in twitter. In Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, 28 March–1 April (2011); pp. 57–58
28. Hoang, T.B.N.; Mothe, J. Predicting information diffusion on Twitter–Analysis of predictive features. J. Comput. Sci.(2018), 28, 257–264.
29. Rao, P.G.; Venkatesha, M.; Kanavalli, A.; Shenoy, P.D.; Venugopal, K. A micromodel to predict message propagation for twitter users. In Proceedings of the 2018 International Conference on Data Science and Engineering (ICDSE), Kochi, India, 7–9 August (2018); pp. 1–5.
30. Naseri, M.; Zamani, H. Analyzing and predicting news popularity in an instant messaging service. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 21–25 July (2019); pp. 1053–1056.
31. S. R. Ahmed, A. K. Ahmed, and S. J. Jwmaa, “Analyzing The Employee Turnover by Using Decision Tree Algorithm,” 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Jun. (2023).
32. N. Z. Mahmood, S. R. Ahmed, A. F. Al-Hayaly, S. Algburi and J. Rasheed, "The Evolution of Administrative Information Systems: Assessing the Revolutionary Impact of Artificial Intelligence," 2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkiye, (2023), pp. 1-7.
33. Peláez, J.I.; Martínez, E.A.; Vargas, L.G. Products and services valuation through unsolicited information from social media. Soft Comput. (2020), 24, 1775–1788.
34. Alduaiji, N.; Datta, A.; Li, J. Influence propagation model for clique-based community detection in social networks. IEEE Trans. Comput. Soc. Syst. (2018), 5, 563–575.
35. Boratto, L.; Carta, S. The rating prediction task in a group recommender system that automatically detects groups: Architectures, algorithms, and performance evaluation. J. Intell. Inf. Syst. (2015), 45, 221–245.
36. L. I. Khalaf, S. A. Aswad, S. R. Ahmed, B. Makki, and M. R. Ahmed, “Survey On Recognition Hand Gesture By Using Data Mining Algorithms,” 2022 International

Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Jun. (2022).

37. L. I. Khalaf, S. A. Aswad, S. R. Ahmed, B. Makki, and M. R. Ahmed, “Survey On Recognition Hand Gesture By Using Data Mining Algorithms,” 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Jun. (2022).
38. B. T. Yaseen, S. Kurnaz, and S. R. Ahmed, “Detecting and Classifying Drug Interaction using Data mining Techniques,” 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Oct. (2022).
39. S. Rashid, A. Ahmed, I. Al Barazanchi, A. Mhana, and H. Rasheed, "Lung cancer classification using data mining and supervised learning algorithms on multidimensional data set," vol. 7, no. 2, pp. 438–447, (2019).
40. Awad, O. F., Sulaiman, S. K., & ALSHMEEL, G. H. A. (2023, October). Anomaly Detection and Security Enhancement Through Machine Learning in Administrative Information Systems. In 2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.
41. Guron, A. T., Anwer, M. A., Sulaiman, S. K., & AbdulSamad, S. J. (2023). Classification of the cause of eye impairment using different kinds of machine learning algorithms. Passer Journal of Basic and Applied Sciences, 5(2), 410-416.
42. Sulaiman, S. K., & Ismael, Y. S. (2022). Machine Learning-Based Personal Signature Recognition Model. NeuroQuantology, 20(10), 2170.
43. Zhou, Y.; Wu, Z.; Zhou, Y.; Hu, M.; Yang, C.; Qin, J. Exploring Popularity Predictability of Online Videos With Fourier Transform. IEEE Access (2019), 7, 41823–41834
44. Chaipornkaew, Piyanuch & Prexawanprasut, Takorn. (2019). A Prediction Model for Human Happiness Using Machine Learning Techniques. 33-37. 10.1109/ICSITech46713.2019.8987513.
45. Aich, Satyabrata & Choi, Ki-Won. (2018). A Machine Learning Approach to Predict Happiness Based on Sentiment Analysis of Twitter Data