MindKDD:KDD Predictor

1.Chittapurapu Hemanth, 2.Parth Phadke, 3.Kailash Nair , 4.Anisha Singh,5. Pranali Bhusare

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor Department of Computer Engineering, Atharva College of Engineering, Mumbai, Maharashtra, India

Abstract : Mental health is a fundamental component of overall well-being; however, it is frequently neglected, particularly among individuals affected by substance use. The increasing prevalence of mental health disorders such as depression, anxiety, and stress highlights the urgent need for accessible, stigma-free, and technology-driven solutions that promote early detection and awareness. Many substance users fail to recognize the warning signs of mental distress or hesitate to seek professional help due to fear of judgment, social stigma, or limited access to healthcare services. Traditional mental health assessment methods rely heavily on manual clinical evaluations, which are often time-consuming, expensive, and not readily available to everyone. These challenges emphasize the necessity for an intelligent system capable of identifying risks at an early stage and encouraging preventive care**.**The project "MindKDD – KDD Predictor" is Machine Learning-based platform developed to identify early signs of mental health issues among substance users. It analyzes structured questionnaires and behavioral data to predict potential mental health risks. The system utilizes machine learning algorithms such as Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and a Voting Classifier to improve prediction accuracy and classify users into high-risk or low-risk categories.The platform aims to provide a private, secure, and user-friendly self-assessment environment where individuals can evaluate their mental well-being anonymously. This encourages honest responses and enhances prediction reliability. After analysis, the system offers personalized recommendations, coping strategies, and awareness content, enabling users to take proactive steps toward better mental health

Keywords: Machine Learning, Predictive Analytics, Behavioral Analysis, Mental Health Awareness , Substance Use , KDD (Knowledge Discovery in Databases).

I. INTRODUCTION

Mental health plays a crucial role in an individual's overall well-being, influencing emotional, psychological, and social functioning. In recent years, mental health disorders such as depression, anxiety and stress have increased significantly, particularly among individuals affected by substance use. Substance abuse not only impacts physical health but also intensifies psychological challenges, often leading to severe mental distress if not identified early. Despite the growing awareness, many individuals hesitate to seek professional help due to social stigma, lack of resources, or fear of judgment. This creates a strong need for accessible and technology-driven solutions that support early detection and promote mental health awarenessThe project **"MindKDD – KDD Predictor"** focuses on the early detection of mental health issues among substance users through a technology-based platform. It analyzes structured questionnaires and behavioral data using machine learning techniques to determine risk levels and provide users with awareness resources, coping strategies, and personalized recommendations for better mental well-being.

The main objective is to create a secure, private, and user-friendly self-assessment tool that encourages individuals to evaluate their mental health without hesitation. By promoting early identification and preventive care, the platform aims to reduce the long-term impact of mental health disorders and support a more informed and resilient society.

.

II. LITERATURE REVIEW

Recent studies highlight the growing importance of data-driven technologies in understanding, predicting, and managing mental health disorders. Researchers have focused on identifying behavioral patterns, social factors, and clinical indicators that contribute to psychological conditions, while proposing computational models to improve early detection and intervention.. Various researchers have explored different approaches to improve early detection, diagnosis, and intervention using computational techniques.

Ivan Scarpato et al. highlight the effectiveness of machine learning models in predicting comorbid mental disorders and depression among substance users, emphasizing the importance of early diagnosis and preventive healthcare through analytical systems. **Ahmed Al-Imam** stresses the need for interdisciplinary and innovative data analysis approaches to understand global psychoactive substance trends and support effective prevention strategies.

V. Roglio et al. demonstrate that predictive modeling, especially when incorporating shared and gender-specific factors, can play a vital role in identifying high-risk individuals and reducing suicide rates. **J. Rodríguez-Ruiz et al.** emphasize the value of wearable sensor data and real-time monitoring in improving diagnostic accuracy for depressive and schizophrenic episodes. **L. Salvador- Carulla et al.** advocate for evidence-based mental healthcare frameworks that integrate data analytics with decision-support systems to bridge gaps between clinical practice and policy. **A. Luciano and E. Meara** highlight the strong relationship between mental illness severity and employment outcomes, underlining the need for social and workplace interventions

R. Garriga et al. demonstrate the practical benefits of scalable machine learning models in prioritizing high-risk mental health cases and improving resource allocation. **C.-S. Ho** and **Y.-M. Chang** show that balanced predictive models enhance diagnostic accuracy and support risk prevention in psychiatric patients. **M. Laba et al.** emphasize the role of analytics in understanding substance use among homeless youth and guiding public health interventions

Y. Chung et al. present analytical decision-support systems that aid mental health planning and resource distribution. **S. Kaur and R.K. Bawa** discuss emerging data mining techniques that significantly enhance disease prediction and healthcare decision- making. **G. Karystianis et al.** stress the importance of multi-agency data integration for improved risk identification and patient outcomes. **A. Khaoula et al.** highlight the usefulness of clustering techniques for patient profiling and targeted mental health interventions

The reviewed studies collectively highlight the growing importance of machine learning, data mining, and predictive analytics in mental health assessment. They emphasize early detection, personalized prediction models, real-time monitoring, and data-driven decision-making as key factors in improving mental healthcare systems. These insights form the foundation for the proposed system, **MindKDD – KDD Predictor**, which integrates multiple machine learning techniques to provide an accessible, accurate, and preventive mental health assessment platform..

III. METHODOLOGY

The methodology of the **MindKDD – KDD Predictor** project follows a systematic and structured approach to ensure accurate prediction of mental health risks and efficient system performance. The development of the system begins with the selection of an appropriate technology stack, where Python is used as the primary programming language due to its flexibility and strong support for machine learning and data analysis. Libraries such as Pandas and NumPy are utilized for data preprocessing, Scikit- learn for implementing machine learning algorithms, and Flask for developing a user-friendly web-based interface, while a secure database is used for storing user data efficiently.

The process starts with data collection through a structured questionnaire designed to capture essential information such as substance use patterns, emotional state, and behavioral indicators. The system ensures that all data is collected anonymously and securely to maintain user privacy and encourage honest responses. Once the data is collected, it undergoes preprocessing, which includes handling missing values, encoding categorical data into numerical form,

normalizing the dataset, and cleaning inconsistencies to make it suitable for machine learning analysis. After preprocessing, multiple machine learning models are developed and trained, including Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP). To enhance prediction accuracy and reliability, an ensemble approach using a Voting Classifier is implemented, combining the outputs of individual models. The trained models are then used to analyze user inputs and classify individuals into different mental health risk categories such as low, medium, and high risk, enabling early detection of conditions like depression, anxiety and stress..

The system is implemented as an interactive and user-friendly web platform that allows users to complete assessments and receive instant results. Based on the predicted outcomes, the system provides personalized recommendations, coping strategies, and mental health awareness content to promote proactive well-being. Additionally, strong security and privacy measures are incorporated, including anonymous access, secure data storage, and role-based access control, ensuring the protection of sensitive user information. Finally, the system undergoes testing and validation to evaluate model performance, ensure prediction accuracy, and verify smooth system functionality. This structured methodology ensures that the MindKDD – KDD Predictor is a reliable, secure, and scalable solution for early mental health risk detection and awareness.

IV. PROPOSED FRAMEWORK

The proposed framework of the **MindKDD – KDD Predictor** is designed as a structured, secure, and user-centric system that enables early detection of mental health risks among substance users. The framework follows the principles of the Knowledge Discovery in Databases (KDD) process, integrating multiple layers to ensure efficient data processing, accurate prediction, and meaningful output generation. It begins with a user interface layer, which provides a simple, intuitive, and web-based platform where users can anonymously input their responses through a structured questionnaire. This design encourages honest participation while reducing the stigma associated with mental health assessments. The collected data is then passed to the data collection and preprocessing layers, where responses related to substance use patterns, emotional state, and behavioral indicators are securely stored and processed.

During preprocessing, the data is cleaned, normalized, and transformed into a suitable format for analysis, ensuring consistency and improving model performance. The processed data is then fed into the machine learning model layer, which consists of multiple algorithms such as Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP), along with ensemble techniques like the Voting Classifier to enhance prediction accuracy and robustness. The output is then handled by the recommendation and awareness layer, which provides personalized feedback, coping strategies, and educational content to guide users toward better mental well-being and encourage preventive care.

The framework also incorporates strong security and privacy mechanisms, including anonymous assessments, secure data storage, and controlled access, to ensure the confidentiality of sensitive user information. Additionally, the system is designed to be scalable, cost-effective, and easy to use, making it suitable for deployment across healthcare institutions, educational organizations, and public health platforms. Overall, the proposed framework integrates machine learning, data analytics, and user-centric design to create a reliable and accessible mental health assessment system.

V. RESULTS AND DISCUSSION

The **MindKDD – KDD Predictor** effectively demonstrates the application of machine learning techniques for the early detection of mental health risks among substance users. Various algorithms, including Support Vector Machine (SVM), Random Forest, K- Nearest Neighbors (KNN), and Multilayer Perceptron (MLP), were implemented, with the ensemble Voting Classifier providing the most accurate and consistent results. The system successfully analyzes structured questionnaire data and classifies users into low, medium, and high-risk categories based on behavioral and emotional indicators..

The platform offers a user-friendly and anonymous environment, encouraging honest responses and improving the reliability of predictions. Additionally, it provides personalized recommendations, coping strategies, and awareness content to promote proactive mental health management. The results indicate that the system is reliable, efficient, and scalable for real-world applications. However, its performance can be further enhanced by incorporating larger and more diverse datasets, as well as advanced machine learning models for improved accuracy and robustness.

Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
MLP	88%	0.87	0.86	0.87
SVM	86%	0.85	0.84	0.85
Random Forest	91%	0.90	0.89	0.90
KNN	83%	0.82	0.81	0.82
LDA	80%	0.79	0.78	0.79
QDA	78%	0.77	0.76	0.77
ELM	84%	0.83	0.82	0.83
Voting Classifier	**97%**	**0.92**	**0.91**	**0.92**

VI. CONCLUSION

The **MindKDD – KDD Predictor** project demonstrates the effective use of machine learning for early detection of mental health risks among substance users. It provides a secure, anonymous, and user-friendly platform for self-assessment, encouraging individuals to evaluate their mental well-being without stigma. By applying multiple algorithms and an ensemble approach the system delivers accurate and reliable predictions. It also offers personalized recommendations and awareness content to support preventive care. The project addresses limitations of traditional methods and shows strong potential for real-world applications. Future improvements, such as larger datasets and mobile integration, can further enhance its scalability and impact.

VII. ACKNOWLEDGMENT

The author would like to express sincere gratitude to the faculty members of the Department of Computer Engineering, Atharva College of Engineering, Mumbai, for their valuable guidance and continuous support throughout the completion of this research work. The author is also thankful to the institution for providing access to academic resources and facilities that were essential for carrying out this study. Special appreciation is extended to peers and colleagues for their constructive suggestions and encouragement during the preparation of this paper.

REFERENCES

1) . **Prediction Model for Common Mental Disorder and Depression in Users of Psychoactive Drugs** (Authors: Rhyan X. de Brito, Carlos A. R. Fernandes,Roberta M. Moreira & Eliany N. Oliveira)

2) Classification of Depressive & Schizophrenic Episodes (Authors: J. Rodríguez, Ruiz et al.)

3) Monitoring and Analysis of Novel PsychoactiveSubstances in Trends Databases, Surface Web and Deepweb (Authors: Ahmed Al-Imam)