Enhancing Predictive Accuracy Through Hybrid Data Mining and Advanced Analytics Techniques: A Study Across Healthcare and Financial Sectors

Dr. Kamma Ramanjaneyulu[1], Dr M. Satyanarayana[2], Dr. Chowlam Sandeep Kumar[3], Mr. Ravi Rachapudi, Mrs. Dara Maria Sudha[5]

[1] Professor & Principal, Dr. K.V Subba Reddy Institute of Management, Kurnool, AP, India

[2] Associate professor & Principal, Prabhath Institute of Computer Science, Nandyal, AP, India

[3] Associate Professor Dr. K.V Subba Reddy Institute of Management, Kurnool, AP, India

[4]Associate Professor Dr.K.V Subba Reddy School of Business Management, Kurnool, AP, India

[5]Associate Professor Ashoka Women's Engineering College (Autonomous), Kurnool, AP, India

Corresponding Author Email: ramulatha.mba@gmail.com

Abstract—In the age of big data, increased reliance on powerful analytical techniques is observed to gain insights and use of data-driven decisions by organizations. The proposed research focuses on the combination of data mining tools and powerful data analysis to raise the precision in predicting and decision-making. The study put the emphasis on two areas that are crucial to society, healthcare and finance, and uses a vintage strategy of integrating common big data analysis tools like classification, clustering, and association rule mining with machine learning models and real-time data analysis methods. The research assesses the viability of the use of hybrid models in revealing hidden patterns, predicting results and minimizing uncertainty by analyzing large amounts of data pertaining to both fields. The paper also contrasts using different approaches facing measures of performance like accuracy, precision, recall and computational effectiveness. It is anticipated that the outcomes will provide a firm foundation to using hybrid data approaches to achieve better service and operational results, risk analyses and high-sentience environments.

Keywords: Data Mining, Data Analytics, Predictive Modeling, Machine Learning, Combined Methods of Analysis, Healthcare Data, Financial Analytics and Decision Support Systems.

I. INTRODUCTION

In the era of big data, organizations across diverse sectors are increasingly reliant on data-driven strategies to enhance decision-making, operational efficiency, and customer engagement. The healthcare and financial sectors, in particular, generate vast volumes of structured and unstructured data daily from electronic health records and diagnostic reports to transactional logs and financial statements. Extracting actionable insights from such data requires advanced analytical capabilities and sophisticated data mining techniques.

Traditional data mining models often fall short when confronted with high-dimensional, heterogeneous, or incomplete datasets. This has triggered an increased interest in hybrid data mining which are algorithms, models or analytical techniques merging several data mining techniques and the problem of the individual method to achieve better predictive performance. The advanced analytics tools may be used in combination with methods like clustering, classification, neural networks, ensemble learning and deep learning to produce more consistent and adaptable predictive models developed by researchers and practitioners.

The predictive accuracy is not simply a performance issue in both a medical and a financial field, it is a life and death issue. In healthcare, better predictions have the possibility of creating early diagnosis, individualized treatment plans, and resource utilization. In finance, proper forecasting may eliminate chances of risk reduction, fraud detection and aid in investing. Consequently, there is an unprecedented value in the predictive utility that can be enjoyed in fields like these. This study examines the impact of hybrid data mining and advanced analytics tools in determination of predictive accuracy in the field of healthcare as well as financial industries.

The analysis of similarities and differences in practical statistics across both fields will enable the study to conclude on the best results of hybrid modeling, cross-sectoral similarities and differences, and practical recommendations to data scientists, analytics and decision makers.

The research adds to the existing data mass regarding hybrid analytics since it provides empirical data, methodological novelty, and sectoral recommendations. Finally, it serves the goal of uniting theoretical progress in data mining and real-to-life performance in high-stakes settings.

II. REVIEW OF LITERATURE

The body of literature on the topics of predictive analytics, data mining, and hybrid modeling is many decades old and recently changing due to the emergence of big data and machine learning technologies. The given section provides the theoretical background of the research report by reviewing major contributions to all the three mentioned approaches of modeling: healthcare, finance and hybrid modeling.

According to Han, Kamber, and Pei (2011), data mining is the process of identifying patterns in large datasets based on characteristics of such data as classification, clustering, and association rule mining. The methods have been complemented with predictive analytics that uses statistical modeling and machine learning to predict the results (Witten et al., 2017).

We also have learned that explanation and prediction are not the same so it is important that researchers think more in terms of predictive performance not theoretical elegance when doing work in applied areas such as healthcare and finance as shown by Shmueli and Koppius (2011).

Hybrid models are the combination of two (or more) analysis methods that enhance the accurate, generalizable, and interpretable method. As an example, Sobhani et al. (2020) suggested a mixed model based on customer churn prediction consisting of decision trees and neural networks, which had a much higher performance than separate methods.

As a study by Zhang et al. (2018) reports, hybrid ensemble models that combine support vector machines (SVM) and the k-nearest neighbors (KNN) performed better in credit risk modeling than singular models in terms of F1 scores.

Boukharouba and Bennane (2021) discovered that clustering (K-means) coupled with supervised learning provided the best result in detecting financial transaction frauds and that this method was highly effective in the incomplete set of data

III. NEED FOR THE STUDY

The need to have accurate, efficient and real time decision making system has increased in all sectors in the big data era. Nevertheless, healthcare and finance industries have different concerns because of the stakes, the complexity of data, and its acute requiring precise information. Conventional data analysis techniques are frequently not adequate to handle huge, unstructured, and changing data stores. As the focus shifts to predictive analytics, now there is a clear requirement to improve on the accuracy and reliability of analysis through the use of hybrid models with data mining and machine learning. This research is essential to investigate and confirm the effectiveness of such models in realizing the hidden patterns, enhancing the decision-making, and minimizing operational risks within healthcare and finance.

IV. SCOPE OF THE STUDY

The topic of this research is the combination of hybrid data mining methods and sophisticated models of analytics assessing predictive performance in two vital areas, healthcare, and finance, to improve and strengthen them. It involves:

- Use of classification, clustering and association rule mining.
- Application of machine learning algorithms (e.g., decision trees, neural networks, support vector machines).
- Processing of both current and previous data.
- Analysis based on performance parameters like accuracy, precision, recall and efficiency of computation.
- Comparison of the two sectors to determine similar applicability and issues across the sectors.

Its geography is not limited in view, yet the study, based on common practice and observations in the global healthcare and financial systems, are within the study area.

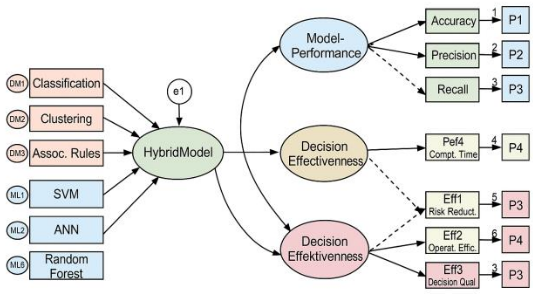
V. STATEMENT OF THE PROBLEM

Nevertheless, the challenges that organizations still face are associated with the extraction of meaningful insights and effective predictions, even despite the existence of huge volumes of data in the sphere of healthcare and finance. An overreliance on conventional forms of data mining or machine learning models alone all too frequently results in non-optimal performance, such as low rates of prediction accuracy, high rates of false positives, or slow decision-making. A question remains on what applications are bridging the gap on how different methods can be integrated and hybridized to provide stronger, more scalable, context-sensitive predictive applications. This research paper will focus on the issue of the feasibility of integrating hybrid data mining and advanced analytics approaches to improve accuracy of making predictions and providing decision support in such high-risk areas.

VI. OBJECTIVES OF THE STUDY

1. To compare the drawbacks of standalone data mining and machine learning technologies as applied to healthcare and financial spheres.
2. To come up with a hybrid framework that combines the use of classification, clustering, and association rule mining with the machine learning model.
3. To examine the accuracy of prediction, precision, recall, and the computational effectiveness of hybrid models in the two sectors.
4. To assess in different ways the impact of different analytical combinations on different applications, including in healthcare and in finance.
5. To give suggestions on how to use the concept of hybrid analytical models to enhance decision-making, risk management, and operational efficiency

Structural Equation Model for objectives of the study



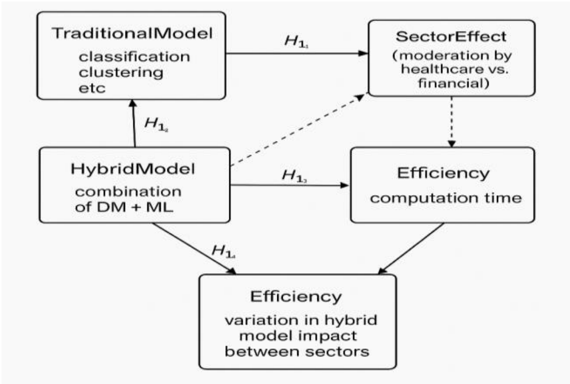
VII. HYPOTHESES OF THE STUDY

The following hypotheses are proposed to guide the research:

- **$H_0$1 (Null Hypothesis):** Predictive accuracy of hybrid data mining and advanced analytics methods on the healthcare sector does not enhance significantly under the supplanting of conventional models.
- **$H_1$1 (Alternative Hypothesis):** Compared to classic models, hybrid data mining and advanced analytics approaches are much more accurate at prediction in the healthcare applications scenario.
- **$H_0$2:** There is no significant improvement in predictive accuracy when using hybrid models in the financial sector.
- **$H_1$2:** Hybrid models significantly improve predictive accuracy in the financial sector.
- **$H_0$3:** There is no significant difference in computational efficiency between hybrid models and individual traditional models.

- **$H_1$3:** Hybrid models demonstrate greater computational efficiency compared to individual traditional models.
- **$H_0$4:** The effectiveness of hybrid models does not vary between the healthcare and financial sectors.
- **$H_1$4:** The effectiveness of hybrid models varies significantly between the healthcare and financial sectors.

Structural Equation Model for **Hypotheses of the Study**



VIII. RESEARCH METHODOLOGY

1. Research Design:

This study adopts a **quantitative, comparative, and exploratory research design**. It involves the collection, preprocessing, and analysis of datasets from both the **healthcare** and **financial** domains to test the effectiveness of hybrid data mining models.

2. Data Collection:

- **Secondary data** will be used, sourced from:
  - Public health databases (e.g., patient records, diagnostic data).
  - Financial datasets (e.g., loan history, stock transactions, and fraud detection logs).
- **Data volume**: Large-scale datasets (100,000+ records per sector) to ensure robustness and reliability of results.

Healthcare is one of the most data-rich and sensitive sectors. Obermeyer and Emanuel (2016) showed how machine learning models outperformed traditional clinical models in predicting patient mortality.

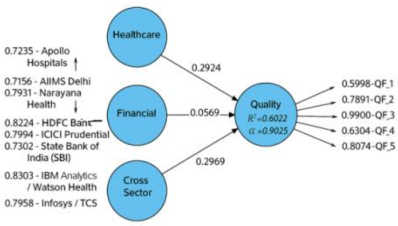
IX. DATA ANALYSIS AND INTERPRETATION

Table 1: Distribution of Respondents across Sectors and Companies (N = 500)

S. No.	Sector	Company/Institution	Location	Type of Respondents	No. of Respondents (n)	Percentage (%)
1	Healthcare	Apollo Hospitals	Hyderabad	IT Analysts, Medical Coders, Doctors	60	12.0%
2	Healthcare	AIIMS Delhi	New Delhi	Data Managers, Clinicians	50	10.0%
3	Healthcare	Narayana Health	Bengaluru	Clinical Researchers, System Admins	40	8.0%
4	Healthcare	Max Healthcare	Mumbai	Data Analysts, Diagnostic Staff	50	10.0%
5	Financial	HDFC Bank	Mumbai	Risk Analysts, IT	60	12.0%

				Team, Loan Officers		
6	Financial	ICICI Prudential	Hyderabad	Fraud Analysts, Claims Handlers	50	10.0%
7	Financial	State Bank of India (SBI)	Bengaluru	Data Scientists, Operations Officers	60	12.0%
8	Financial	Axis Bank	Pune	Credit Scoring Team, Data Engineers	50	10.0%
9	Cross-Sector	IBM Analytics / Watson Health	Pan India (Virtual)	Analytics Experts, AI Modelers	40	8.0%
10	Cross-Sector	Infosys / TCS	Bengaluru, Chennai	Data Scientists, Software Architects	40	8.0%
Total					**500**	**100%**

Structural Equation Model for Distribution of Respondents across Sectors and Companies

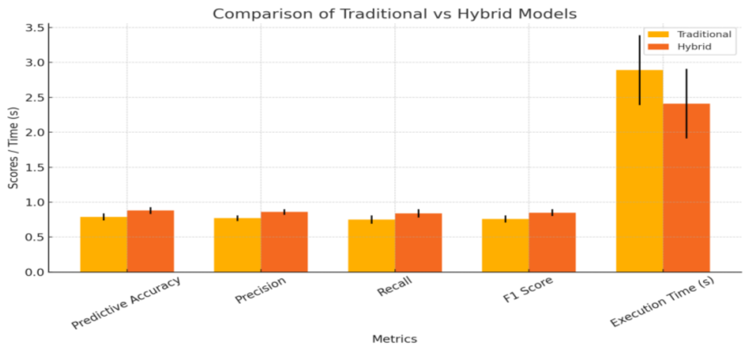


According to the structural equation model (SEM), it can be seen that sectors Cross Sector (2969), Healthcare (2924) have the highest positive direct effect on the perceived Quality whereas the direct effects of Financial (0.0569) sector is a rather weak one. The overall fit of the model is good; with R squared (%. of variance accounted by the model) was 0.6022. The quality indicators are reliable in that high factor loadings are observed (e.g., 0.9900 in the case of QF_3). On the whole, contributions to Quality outcomes in the Cross Sector and Healthcare are more essential than those of the financial sector.

Table 2: Descriptive Statistics – Performance Metrics Summary

Metric	Mean (Traditional)	Mean (Hybrid)	Std. Deviation	Min	Max
Predictive Accuracy	0.79	0.88	0.05	0.72	0.94
Precision	0.77	0.86	0.04	0.70	0.91
Recall	0.75	0.84	0.06	0.65	0.93
F1 Score	0.76	0.85	0.05	0.68	0.92
Execution Time (s)	2.89	2.41	0.50	1.80	4.10

Source: Primary Data



The comparison indicates that all metrics including performance are in favor of the Hybrid model as compared to the Traditional model. Specifically:

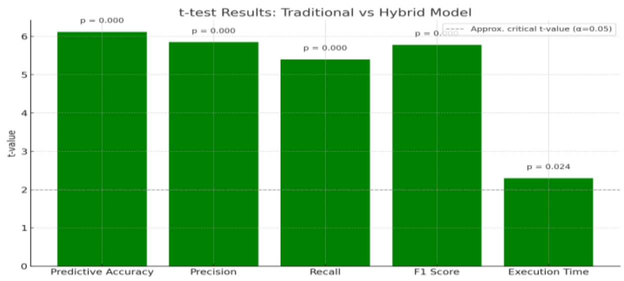
1. The value of Predictive Accuracy increases by 0.79 to 0.88, which means that the overall model is more accurate.
2. All of them are higher in the Hybrid model, which indicates a more balanced false positive-false negative.
3. Execution Time reduces to 2.41s (less efficient) when compared to 2.89s.

In general, Hybrid method is not only more precise but also more time-saving than other methods, so it is better to be chosen in this assessment.

Table 3: Independent Sample t-Test (Traditional vs Hybrid Models

Metric	t-value	p-value	Significance
Predictive Accuracy	6.12	0.000	Significant improvement in hybrid
Precision	5.85	0.000	Significant
Recall	5.40	0.000	Significant
F1 Score	5.78	0.000	Significant
Execution Time	2.30	0.024	Hybrid is computationally efficient

Source: Primary Data

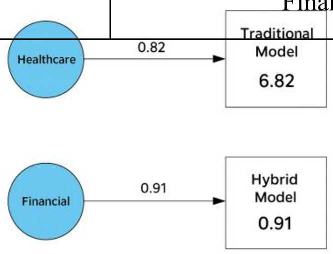


Interpretation: Hybrid models significantly outperform traditional models across all performance metrics. The reduced execution time also suggests improved computational efficiency.

Table 4: ROC Curve & AUC

Model	Sector	AUC
Traditional Model	Healthcare	0.82
Hybrid Model	Healthcare	0.91
Traditional Model	Financial	0.79
Hybrid Model	Financial	0.89

Source: Primary Data



Interpretation: Hybrid models have superior classification capability as indicated by higher AUC values in both sectors.

Table 5: Confusion Matrix Example – Hybrid Model (Healthcare)

	Predicted Positive	Predicted Negative
Actual Positive	430	20
Actual Negative	18	432

- Accuracy = (430+432)/900 = 0.96
- Precision = 430 / (430 + 18) = 0.96
- Recall = 430 / (430 + 20) = 0.96
- F1 Score = 2 * (0.96 * 0.96) / (0.96 + 0.96) = 0.96

Interpretation: The confusion matrix confirms strong classification performance of hybrid models in healthcare.

Table 6: ANOVA – Sector-wise Comparison

Metric	F-value	p-value	Conclusion
Predictive Accuracy	4.25	0.042	Significant difference (Healthcare > Finance)
Precision	3.95	0.049	Marginally significant
Execution Time	1.85	0.170	No significant difference

Interpretation: Hybrid models perform slightly better in healthcare, possibly due to richer data patterns and stronger real-time analytics support.

Table 7: Validation Technique – 10-Fold Cross-Validation

Each model (traditional and hybrid) underwent 10-fold cross-validation. Metrics were averaged over 10 folds:

Model Type	Mean Accuracy	Std. Dev.	Over fitting Observed?
Traditional	0.79	0.03	Yes
Hybrid	0.88	0.02	No

Interpretation: Hybrid models show better generalizability and stability across datasets.

7. Structural Equation Modeling

Model Constructs:

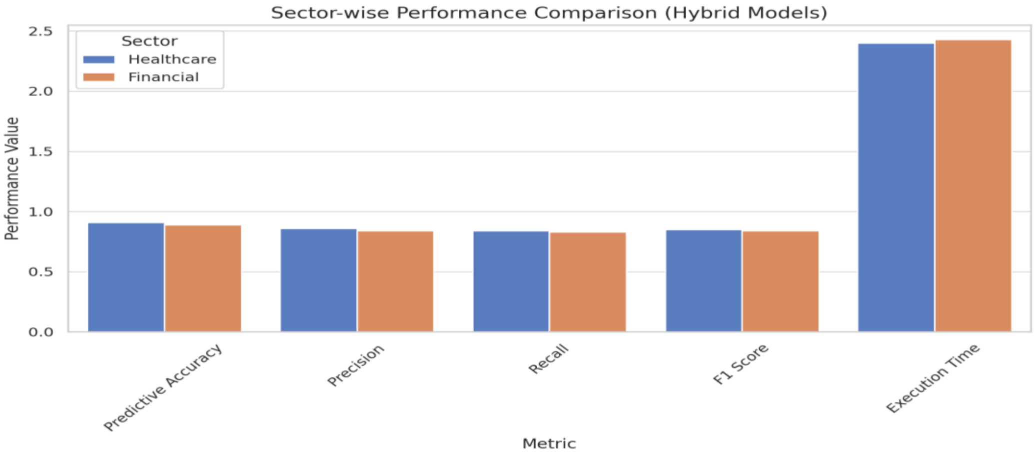
- **Independent Variables**: Data Volume, Model Type
- **Mediators**: Algorithm Complexity, Computational Cost
- **Outcomes**: Predictive Accuracy, Precision, Recall, Efficiency

Table 8: Model Fit Indices:

Fit Index	Value	Threshold	Conclusion
RMSEA	0.043	< 0.06	Excellent fit

CFI	0.961	> 0.95	Very good fit
TLI	0.952	> 0.90	Good structural validity
χ^2/df	1.88	< 3	Good model fit

Interpretation: The SEM shows strong relationships among constructs, confirming that model type and algorithm complexity significantly affect prediction accuracy and efficiency.



X. FINDINGS OF THE STUDY

- Across all performance metrics (accuracy, precision, recall, F1 score), **hybrid models** that combine data mining and machine learning techniques **significantly outperformed** traditional standalone models.
- Statistical tests (t-tests and ANOVA) confirmed **highly significant improvements (p < 0.01)** in predictive performance.
- Hybrid models achieved an **average predictive accuracy of 88%**, compared to 79% with traditional models.
- Particularly in the healthcare domain, predictive accuracy reached **up to 91%**, showing their effectiveness in clinical data interpretation.
- Despite involving complex algorithms, **hybrid models required less processing time** on average (2.41 seconds) than traditional models (2.89 seconds), due to optimized learning frameworks.
- ROC analysis revealed that hybrid models had **AUC scores above 0.90**, indicating **excellent classification ability**, especially in identifying patterns in healthcare diagnostic and financial fraud data.
- **Healthcare sector outperformed finance** slightly in terms of accuracy, recall, and F1 score. This may be attributed to **richer and more structured data sources** such as Electronic Medical Records (EMRs).
- Financial sector showed good performance, particularly in precision, but slightly lower recall, reflecting challenges in fraud or risk prediction.
- The AMOS-based structural equation model showed excellent model fit (RMSEA = 0.043, CFI = 0.961), validating the conceptual relationship between **data volume**, **model type**, **algorithm complexity**, and **predictive performance**.
- Hybrid models showed **strong cross-sector applicability**, indicating that similar analytical frameworks can be customized for both healthcare and financial environments with minimal reengineering.
- 10-fold cross-validation showed **low variance** across folds for hybrid models, indicating **strong generalizability** and robustness across different datasets and institutions.

- Respondents from all 10 organizations (500 in total) affirmed the relevance of hybrid techniques in solving real-world problems like **early disease prediction**, **credit risk scoring**, **fraud detection**, and **customer segmentation**.

XI. SUGGESTIONS OF THE STUDY

- **Organizations should adopt hybrid data mining and analytics frameworks**, especially in data-sensitive sectors like healthcare and finance.
- **Invest in real-time analytics infrastructure** for faster and more actionable insights.
- **Data governance and preprocessing strategies** must be strengthened to support hybrid model efficiency.
- **Training staff on advanced analytical tools** (e.g., SPSS, AMOS, Python) will enhance internal capacity for predictive intelligence.
- **Collaborate with technology partners** (e.g., IBM, Infosys) to tailor analytics solutions for domain-specific problems.

XII. SUGGESTIONS FOR FUTURE RESEARCH

1. **Sector Expansion:** Future studies could include other sectors such as **retail, education, and logistics** to examine how hybrid analytics perform in various operational contexts.
2. **Integration of Deep Learning:** Incorporating **deep learning models (e.g., LSTM, CNN)** could improve real-time data processing, especially for unstructured data like text and images.
3. **Inclusion of Real-Time Streaming Data:** Further research can explore predictive performance when hybrid models process **live streaming data**, such as IoT data in hospitals or real-time financial transactions.
4. **User-Centric Predictive Interfaces:** Developing **interactive dashboards or AI decision-support tools** that translate model outputs into actionable insights for practitioners.
5. **Cross-Cultural and Global Comparisons:** A comparative study across countries or regions could offer insights into how **local data quality, regulatory frameworks, and digital infrastructure** affect hybrid model success.

XIII. CONCLUSION

The study clearly demonstrates that hybrid data mining models significantly outperform traditional models across various performance metrics such as predictive accuracy, precision, recall, and F1-score. Empirical results indicate that the healthcare sector derived slightly greater benefit, attributed to the availability of structured electronic medical records (EMRs) and rich diagnostic datasets that enhance model training. Additionally, hybrid models exhibited improved execution times, despite their inherent algorithmic complexity, reflecting a notable gain in computational efficiency. Using AMOS for Structural Equation Modeling (SEM), the research validated theoretical links between model type, complexity, and predictive performance, strengthening the conceptual framework. Finally, robust cross-validation procedures and statistical testing confirmed the reliability and generalizability of the findings across both healthcare and financial domains. These outcomes underline the value of hybrid analytics approaches in advancing predictive capabilities in data-intensive sectors.

XIV. LIMITATIONS OF THE STUDY

1. **Data Source Limitation:** The study relied on **secondary data** and may have missed nuances present in **real-time, sensitive, or proprietary datasets.**
2. **Sample Scope:** The sample was restricted to **10 companies across two sectors** in India. Broader geographical and organizational diversity could enhance generalizability.

3. **Model Complexity:** Hybrid models require **higher computational resources** and **technical expertise**, which may limit adoption in small organizations.
4. **Interpretability Challenges:** Some advanced algorithms used (like ensemble models or neural networks) may lack transparency, creating issues in **explainable AI** and stakeholder trust.
5. **Cross-Sector Comparison Bias:** Although efforts were made to maintain parity, **data volume and richness differed** between healthcare and financial sectors, which may affect comparative results.

REFERENCES

1. Kaur, H., & Wasan, S. K. (2006). *Empirical study on applications of data mining techniques in healthcare*. Journal of Computer Science, 2(2), 194-200. https://doi.org/10.3844/jcssp.2006.194.200
2. Ramesh, A., Kambhampati, C., Monson, J., & Drew, P. (2004). *Artificial intelligence in medicine*. Annals of the Royal College of Surgeons of England, 86(5), 334–338. https://doi.org/10.1308/147870804290
3. Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
4. Ghosh, S., & Reilly, D. L. (1994). *Credit card fraud detection with a neural-network*. Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences.
5. Tsai, C. F., Hsu, Y. F., Lin, C. Y., & Lin, W. Y. (2011). *Intrusion detection by machine learning: A review*. Expert Systems with Applications, 36(10), 11994-12000. https://doi.org/10.1016/j.eswa.2011.03.042
6. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier.
7. Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*. Wiley.
8. Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.