WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Visual Insights: A Comprehensive Approach to Pre-Processing for Analytics

***Ayush Garg[1], Dr. Vishal Shrivastava[2], Dr. Akhil Pandey[3], Dr. Karuna Sharma[4], Dr. Ashok Kumar Kajla[5]***

[1]B.Tech. Scholar, [2-5]Professor

Department of Artificial Intelligence & Data Science, Arya College of Engineering and I.T. Jaipur, India,

ayushgokul2004@gmail.com[1], vishalshrivastava.cs@aryacollege.in[2], akhil@aryacollege.in3,skaruna.cs@aryacollege.in[4], ashokkajla@aryacollege.in[5]

**ABSTRACT**

Data visualization and pre-processing are fundamental pillars of data analytics, transforming raw datasets into meaningful insights. This research explores the intertwined roles of data pre-processing and visualization in modern data science workflows. Pre-processing, encompassing data cleaning, transformation, and normalization, is essential for addressing inconsistencies, handling missing values, and preparing datasets for analysis. Visualization, on the other hand, offers intuitive representations of complex data, enabling pattern recognition, trend analysis, and informed decision-making.

This paper examines the methodologies, tools, and challenges associated with data pre-processing and visualization. It highlights the importance of pre-processing in ensuring data quality and integrity, which directly influences the accuracy and interpretability of visualizations. Furthermore, it investigates the impact of emerging technologies such as artificial intelligence and machine learning in automating pre-processing tasks and generating dynamic, interactive visualizations.

Case studies and real-world applications are presented to illustrate how integrated pre-processing and visualization workflows can uncover hidden patterns and foster actionable insights. The research also delves into common pitfalls, such as biases introduced during data preparation and misinterpretation of visualizations, and offers strategies to mitigate them.

By bridging the gap between these two critical stages of data analysis, this paper provides a comprehensive framework for researchers and practitioners to enhance the efficacy of their data-driven initiatives.

KEYWORDS: Data Pre-Processing, Data Visualization, Data Analytics, Data Quality, Data Cleaning, Transformation , Pattern Recognition, Interactive, Visualizations, Machine Learning, Emerging Technologies, Actionable Insights, Bias Mitigation, Workflow Integration, Data Interpretation, Trends and Patterns, Big Data Analytics, Data Visualization, Association Rule Mining Knowledge Mining, Visualization

## 1. INTRODUCTION

Organizations, researchers, and decision-makers are now faced with the volumes and varieties of data that have reached unprecedented high velocities. Raw data has to be culled for meaningful insights. However, this is an opportunity on the other side and a challenge on one side. Sufficient methodologies are required so that proper information processing and presentation is achieved. This includes two related activities: data pre-processing and data visualization. Together, they make up the bedrock of the modern data analytics pipeline in ensuring that data-driven decisions are accurate, reliable, and actionable.

### *1.1 Data Pre-Processing*

Data pre-processing is the first critical step in the data analysis lifecycle. It prepares raw data for analysis by addressing issues of missing values, inconsistencies, noise, and outliers. Key tasks in pre-processing include data cleaning, transformation, normalization, and feature engineering. Improved quality and consistency in data reduce errors and biases and makes subsequent analyses stand on firmer ground.

The preprocessing stage is most critical when dealing with heterogeneous datasets in which data originates from more than one source, thus having different formats and structures. Without such preprocessing, analyses may be faulty, leading to misleading conclusions. It's even more significant in sensitive domains like health, finance, and climate science because it impacts the accuracy of insights profoundly.

While pre-processing handles the quality of data, visualization handles its interpretability. Data visualization transforms processed datasets into graphical representations, letting users identify patterns, trends, and anomalies in record time. Good visualizations break down complex datasets into forms that can be intuitively understood and enhance understanding in a manner that enables people to use data to narrate their stories. Visualization is not an end product of analysis but a dynamic tool which can be used in exploratory data analysis. Very early in the analysis process, visualizing data can lead to the discovery of hidden relationships and patterns that would guide further exploration. State-of-the-art visualization methods, such as interactive dashboards and 3D plots, allow users to manipulate and explore data in real time, thereby inducing deeper engagement and insight.

### *1.2 The Interplay Between Pre-Processing and Visualization*

There is a strong relationship between good quality of pre-processing and effective data visualization; pre-processing that is shoddily done can introduce misrepresentations in visualization. So, in turn, insight into visualizations provides one of the ways to discern that some pre-processing would have been required, be it detection of outliers or anomaly identification. This interplay has become stronger with the increase in the use of automation and ML applications in data workflows. Many pre-processing jobs are now automated by algorithms while advanced visualization tools deliver real-time, interactive graphics. However, these advancements introduce a new set of problems with over-reliance on algorithms and a call for transparency in processes automated.

### *1.3 Research Motivation and Objectives*

This research is about the synergistic relationship between data pre-processing and visualization, in the context of their role in improving the effectiveness of the analysis. It will:

1. Investigate the methods and tools used in data pre-processing and visualization.

2. Examine the effect of pre-processing on the accuracy and interpretation of visualization.

3. Identify recurring issues and provide recommendations on how to reduce biases and misinterpretations.

4. Discuss real-world applications to demonstrate the relevance of combining these processes in practical life.

### *1.4 Paper Outline*

The next parts introduce theoretical concepts and practical applications concerning preprocessing and visualization of data. This section reviews recent developments in literature. Then methodology presents the analytical approach used in this study. Cases are practical applications, which, in the following discussion, will present issues and challenges, limitations and limitations of the study, and also future research prospects.By addressing these aspects, this paper aims to provide an all-inclusive understanding of how data pre-processing and visualization can be effectively integrated in order to maximize the value of data analytics.

## 2. Literature Review

Data visualization and data pre-processing techniques have been developed together with an increase in the complexity of datasets and the need for even more complex analytical tools. This section traces the theoretical background, recent advances, as well as challenges involved in two related fields.

### *2.1 Evolution of Data Pre-processing Techniques*

Data pre-processing has always been an integral part of data analysis. Its origin was at the beginning of the earliest statistical methods used in cleaning and preparing the data. With advancements in computational technologies over time, more complex kinds of pre-processing have found their place. Some examples are as follows:

***Data Cleaning and Transformation***: Outlier, inconsistency, or missing values have matured in methods to handle, with imputation algorithms or robust statistical techniques ensuring data quality.

***Feature Engineering and Selection:*** Techniques such as PCA, one-hot encoding, and feature scaling have become routine to make datasets relevant and interpretable for machine learning models.

***Automated Pre-Processing***: Thanks to tools and frameworks such as TensorFlow Data Validation and AutoML, automated pre-processing workflows are now possible and reduce the effort that needs to be invested in manual work with higher consistency.

Poor preprocessing can lead to biased analyses and spurious conclusions according to research; thus, it forms an essential step in data-driven research.

### *2.2 Innovation in Data Visualization*

Data visualization started from static charts and graphs to dynamic interactive systems through which users can explore and analyze data in real-time. Some of the significant innovation in this field includes:

***Interactive Visualization Tools***: The advanced methods for visualizing data through tools such as Tableau, Power BI, and D3.js, which have made the creation of dashboards and reports quite easy without much technical background.

***Visual Analytics***: When computation was combined with the visualization, it resulted in a new form of analytics called visual analytics. With this, large data bases are explored interactively. This way, hidden patterns are found.

***Storytelling in Visualization***: It is now highly emphasized that clear context with an audience-specific design needs to be followed, as highlighted by contemporary research about effective storytelling in visualization.

While great strides have been taken toward understanding data visualization, the issue of misrepresentation and cognitive bias remains. Therefore, more study into best practices is an inevitable task.

### *2.3 Pre-Processing Interdependencies*

There is a really symbiotic interplay between the pre-processing and visualization activities. Good visualizations must be based on clean and well-prepared data, while the visualization can turn up the insights that help shape even further pre-processing. Key points that have emerged through the literature in this direction include the following:

***Feedback loops***: Visualization helps find abnormalities, missing data and inconsistencies that need even further pre-processing. The techniques to reduce the dimensionality, be it t-SNE or PCA, which are generally done at the pre-processing level, affect the quality and interpretability of visualization in high dimensional datasets.

***Bias Detection:*** Visualization tools enable detection of biases in the data and taking remedial action at the preprocessing stage.

This work has been fundamentally to provide an integrated workflow that will close this gap between the two; namely, bridge pre-processing with visualization, toward realizing a more iterative and exploratory process for data analysis.

### *2.4 Emerging Technologies and Trends*

The rapid acceleration of artificial intelligence and machine learning is enabling new paradigms, both in the space of pre-processing and in the realm of visualization. Some trends include:

**AI-aided Pre-Processing:** Techniques such as anomaly detection using deep learning, automated feature selection are the game-changers in pre-processing of raw data for analysis.

**Augmented Analytics**: Combination of AI with visualization tool allows real-time insights, automated chart recommendations, and predictive analytics.

**Big Data Visualization:** Big data calls for new forms of visualizations that can handle big and heterogeneous datasets, such as heatmaps, network graphs, geospatial plots. Despite these developments, the challenges remain high, such as scalability of algorithms, interpretability of AI models, and ethical issues related to data privacy and bias.

### *2.5 Gaps in Current Literature*

Much has been achieved despite which gaps still exist in the literature:There is little work that has addressed the integration of automated pre-processing and visualization workflows.

There is little exploration that has been undertaken into the ethical considerations surrounding AI-driven pre-processing and visualization.

A demand for standard guidelines to remove biases and increase the interpretability of visualizations.

## 3. Methodology

This research is focused on the technological frameworks, tools, and processes that underlie successful data pre-processing and visualization integration. It is divided into three phases: data acquisition and preparation, pre-processing, and visualization. Each of the phases utilizes advanced technologies in ensuring that there is full and accurate analysis.

### *3.1 Data Acquisition and Preparation*

It starts with gathering datasets from different sources, starting with public repositories to APIs and case studies about real-life projects. Among the technological tools and techniques used in this process is the following:

***Data Sources:*** Sourced datasets from various websites such as Kaggle, UCI Machine Learning Repository, and government open data portals.

***Data Collection Tools:*** Applied Python libraries, including requests, pandas, and BeautifulSoup, in collecting data by either scrapping or APIs.

***Data Storage:*** Utilized cloud-based storage systems, like AWS S3 and Google **BigQuery**, for the large amount of data. Therefore, it is scalable and accessible.

Data Pre-Processing

Pre-processing is used to ensure the raw data is clean and consistent and is ready for visualization. Technologies and techniques applied:

***Data Cleaning:***

Filled missing values using imputation techniques by scikit-learn.

Identified and removed outliers with statistical methods and visualization tools, like box plots.

***Data Transformation:***

Applied feature scaling using NumPy and pandas for min-max normalization and standardization.

Categorical variable encoding by one-hot encoding and label encoding by scikit-learn

***Dimensionality Reduction:***

The high-dimensional data is reduced using PCA and t-SNE with both scikit-learn and TensorFlow

Automated Pre-Processing

**AutoML** platforms including Google **AutoML** and H2O.ai have been used to automate the boring pre-processing tasks.

All the above pre-processing workflow was done using **Jupyter** Notebooks to give transparence, reproducibility, and ease of collaboration.

### *3.2 Data Visualization*

The visualization phase is to create meaningful, intuitive, and interactive representations of the processed data. Advanced visualization technologies and frameworks were used in this phase, including:

***Static Visualizations***

Creation of charts and plots such as line graphs, histograms, scatter plots using Matplotlib and Seaborn.

***Interactive Dashboards***

Development of dynamic dashboards using tools such as Tableau and Python-based frameworks like **Plotly** Dash and **Streamlit**.

Real-time data streams integration for interactive exploration.

***Big Data Visualization***:

Worked with big data libraries and visualization tools optimized for big data, for instance, Apache Superset, **ECharts**, etc., to work with large datasets

Used Spark-based tools that offer distributed visualization of huge datasets

***Geospatial Visualization***

Applied folium and **GeoPandas** in the spatially mapping of data and overlaying information layers.

### *3.3 Integration and Workflow Automation*

The research enhanced efficiency and scalability through the integration of workflows and automation

***ETL Pipelines:***

Built an ETL pipeline using Airflow for seamless switching between pre-processing and visualization.

***API Integration:***

Integrated data APIs with visualization dashboards with the help of Flask and **FastAPI** for real-time updation.

***Cloud-Based Platforms:***

Deployed pre-processing scripts and dashboards on Google Cloud Platform, and Microsoft Azure to enjoy the advantage of remote accessibility and scalability.

### *3.4 Evaluation Metrics*

The following metrics were used to judge the methods:

***Pre-processing Accuracy***: Measured in terms of missing data reduction, consistency of datasets, and performance enhancement of the model after preprocessing.

***Effectiveness of Visualization:*** Based on user feedback about interpretableness, clarity, and usability of the visualizations generated.

***Scalability and Performance:*** This method was tested in terms of complexity of processing and visualization by its pipelines over different sizes of data sets.

| **Phase** | **Tools and Technologies** |
|---|---|
| Data Acquisition | Python (requests, pandas, BeautifulSoup), APIs |
| Data Pre-Processing | scikit-learn, NumPy, pandas, TensorFlow, AutoML |
| Visualization | Tableau, Matplotlib, Seaborn, Plotly Dash, Streamlit |
| Workflow Integration | Apache Airflow, Flask, Google Cloud Platform |

## 4. Case Studies and Applications

This section highlights real-world applications of data pre-processing and visualization, illustrating their importance across various domains. Each case study demonstrates how these processes enable meaningful insights and informed decision-making.

### *4.1 Healthcare Analytics*

***Overview***

In the healthcare industry, data pre-processing and visualization are critical for handling large, heterogeneous datasets such as patient records, medical imaging, and genomic data.

***Case Study***

There is a hospital system that made used of pre-processing techniques in their mining of patient admission data. Some critical steps followed were: They ensured cleaning of inconsistent EHR formats. Normalizing lab test results was included because they were on other scales. The process used PCA to reduce the dimension of patient features prior to clustering. It visualized the preprocessed data

- Trend pattern of patient admission rates along seasons using heatmaps

- Correlation between various demographic factors and common diseases using scatter plots

- Interactive dashboards to indicate real-time ICU occupancy

**Impact**

The integration of pre-processing and visualization improved resource utilization, reduced waiting times among patients, and enhanced strategic planning.

### *4.2 Financial Data Analysis*

**Summary**

The financial sector generates massive volumes of transactional and market data, requiring stringent preprocessing and effective visualization for actionable insights to be derived.

**Case Study**

A fintech firm analyzed the trends in the stock market by integrating historical trading data from various exchanges. The critical steps undertaken included:

- During preprocessing, differences in time zones and currency were accounted for.

- Outlier detection was implemented to flag incorrect trades.

- Application of rolling averages and feature engineering to extract key indicators.

**Data visualizations included:**

Interactive candlestick charts of the movement of stock prices, network graphs showing interrelation among market sectors, and dashboards with predictive analytics for portfolio management.

**Impact**

The company improved the accuracy in the forecasting of market trends and offered the clients better decision-support tools.

### *4.3 Environmental Monitoring and Climate Data*

**Overview**

Environmental science relies on pre-processing and visualization to make sense of complex datasets, such as weather patterns, pollution levels, and biodiversity metrics. Case Study Scientists used satellite data to assess the trend of air pollution in urban areas. In this pre-processing stage, filtering and cleaning raw satellite data with noises were done.

Data summarization across spatial grids to make it uniform. Normalization of pollutant concentration to analyze the temporal aspect.

**Visualizations include:**

- Geospatial maps of hotspots of pollution

- Time-series charts showing pollutant levels over the last few decades

- Dashboards comparing pollutant levels in cities

**Impact**

The study would help policymakers pinpoint the zones of high risk and undertake targeted environmental policies.

### *4.4 Business Intelligence and Decision Support Systems*

**Overview**

Businesses employ data pre-processing and visualization to push the optimization of operations, better customer experiences, and profitability.

**Case Study**

The sales data from a chain of retail stores was analyzed as part of the retail chain's efforts to adjust its inventory management to improved levels. Pre-processing activity included the following steps to handle missing sales records as well as standardize categories of products.

- Clustering of stores based on sales performance.

- Inclusion of external data, such as weather and holidays, to enhance richness in the insights.

In terms of visualizations:

- Sales heat maps across geographic regions.

- Bar charts for monthly revenues by category.

- Dashboards that predict inventory need based on seasonal trends.

**Impact**

This approach reduced overstocking and stockouts while increasing operational efficiency and improving customer satisfaction.

Common trends that arose from these case studies were:

1. Interdisciplinary Integration: For successful application, it has to happen in collaboration with domain experts and data analysts.

2. Dynamic Workflows: Iterative feedback loops between pre-processing and visualization yield better insights of data

3. Scalable Solutions: Tools and techniques used should address increasing data volume without losing performance.

All these case studies underpin the practical relevance of data pre-processing and visualization; they are transformational forces to solve real-world problems.

## 5. Conclusion and Future Directions

Data pre-processing and visualization are merged to extract meaningful insights from complex data sets. Using this study, we showed the interrelated roles that these two processes play, and assessed their importance in improving quality and interpretability of the data analysis process. Modern technologies were thus used to show how data pre-processing ensures clean, consistent, and well-prepared data, while data visualization promotes intuitive and impactful communication of findings.

### *5.1 Summary of Findings*

This paper has a few critical conclusions on data pre-processing and visualization as below:

**Data Pre-Processing is Necessary**: Any data processing or preparation directly affects the integrity and quality of data. Unless effectively pre-processed, analysis or further visualization might not be trustworthy. Important activities such as cleaning of data, transformation, and normalization enhance the quality of the data to make it more user-friendly for analytical purposes.

**Visualization Helps Complex Data Be More Readable**: One of the key aspects in helping complex data become more readable is visualization. It makes raw data appear in a more aesthetically pleasing form and points out the existence of patterns, trends, or anomalies. More sophisticated visualizations, like interactive dashboards and geospatial visualizations, have turned to be particularly valuable in health care, finance, and environmental monitoring applications.

**Symbiotic Relationship**: Data pre-processing and data visualization do not operate independently but together; well-preprocessed data tends to often provide clearer, more accurate visualizations and similarly, good visualizations might often point to areas requiring more pre-processing to uncover outliers or missing data, for example.

**Emerging technologies drive innovation**: Machine learning (ML), artificial intelligence (AI), and automation have transformed the pre-processing and visualization processes. Automated feature engineering, AI-based anomaly detection, and interactive real-time visualization have improved the efficiency and scalability of these processes.

### *5.2 Recommendations for Practitioners and Researchers*

Based on findings, a number of recommendations for practitioners and researchers are proposed here:

**Emphasize Quality Pre-Processing:** Practitioners should emphasize thorough pre-processing in their data workflows. Tools that automate data cleaning, feature engineering, and transformation can significantly reduce human error and improve the efficiency of the analysis pipeline. Ensuring that data is normalized and standardized across datasets will improve the interpretability of visualizations.

**Use Interactive and Real-Time Visualizations**:

Interactive visualizations, such as dashboards and dynamic charts are to be used in order to help researchers and organizations make deeper engagement with data. The use of these tools in real-time allows data exploration, hence allowing a decision-maker to manipulate these visualizations for insights of customized information.

**Improve Interdisciplinary Collaboration:** Effective integration of pre-processing and visualization calls for effective collaboration between data scientists, domain experts, and decision-makers. It ensures that the methods and visualizations selected are customized to the needs and peculiarities of the field.

**Focus on Ethical Considerations**: Practitioners and researchers should be ever vigilant of ethical concerns while processing and visualizing data. Bias in data, at the pre-processing stage or through selective visualization choices, can guide decision-making processes astray. Bias detection mechanisms and transparency through automated processes are essential to establishing trustworthiness in data-driven decisions.

**Leverage Cloud-Based Tools for Scalability**: As data sizes grow, managing and visualizing massive datasets efficiently needs to be handled by leveraging the cloud platforms of AWS, Google Cloud, or Microsoft Azure. They provide scalable storage, computing power, and integration with visualization tools to handle complex, high-dimensional datasets.

### *5.3 Potential Areas for Future Research*

Though this paper provides an all-inclusive overview of data pre-processing and visualization, still many areas remain open to further investigation:

**Automated Pre-processing of Advanced Complex Data Types**: There will always be future research in advance automation tools to preprocess such unstructured data of the texts, images, or audio. The machine algorithms will automatically recognize and change the noisy and irrelevant ones with minimal intervention of man.

The explanation of machine learning models, where models are being applied increasingly to data analysis, has brought about the requirement to increase interpretability in such models used within visualizations. Introducing and investigating ways in which results of machine learning can be incorporated into tools that visualize and interact with analytical systems will build confidence in insights coming from AI.

**Big Data Visualization**: Big data comes with certain visualization challenges, especially on the real-time processing of huge amounts of data. Future studies may focus on the new visualization methods that can process and present big data in an intuitive manner, including real-time stream processing visualization or distributed visualization platforms.

Visualization and Ethical Considerations: The use of visualizations in decision-making contexts requires further research to be conducted on the ethical aspects. Future work may look at the biases in the selection of visualizations, say misleading chart types or selectively represented data, and discuss how these can be eliminated.

**Future data science workflows**: In the near future, end-to-end platforms would be fully integrated and automated, encompassing all steps of pre-processing, analysis, and visualization in one system. To realize this dream of making data analysis accessible and efficient, large-scale research into the best ways to integrate these pieces will be needed—user experience, scalability, and real-time interactivity will come out on top.

## Conclusion

In conclusion, data pre-processing and visualization are core parts of the pipeline of data analysis; actually, these influence accuracy, efficiency, and the ease of interpreting insights. Indeed, as advanced technologies and methodologies evolved, so did pre-processing and visualization and allowed for more powerful, scalable solutions for both of them. However, some issues remain and exciting research directions include areas such as automation in pre-processing, the visualization of big data, and ethical considerations. This means that the interaction between pre-processing and visualization will get more robust with advancing technology and pave the way for effective, insightful, and responsible data analysis.

### REFERENCES

[1] K. Shailaja, B. Seetharamulu, M. A. Jabbar, “Machine Learning in Healthcare: A Review”, 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), DOI: 10.1109/ICECA.2018.8474918.

[2] Aishwarya Mujumdar, V. Vaidehi, “Diabetes Prediction using Machine Learning Algorithms”, INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019, DOI: 10.1016/j.procs.2020.01.047.

[3] Aiswarya Iyer, Dr. S Jeyalatha, Ronak Sumbaly, “Diagnosis of Diabetes Using Classification Mining Techniques”, International Journal of Data Mining and Knowledge Management Process (IJDKP) Vol.5, No.1, January 2015, DOI: 10.5121/ijdkp.2015.5101.

[4] Talha Alam, Muhammad Iqbal, Yasir Ali, Abdul Wahab, Safdar Ijaz, Talha Baig, Ayaz Hussain, Muhammad Malik, Muhammad Raza, Salman Ibrar, Zunish Abbas, “A model for early prediction of diabetes”, January 2019, Informatics in Medicine Unlocked 16:100204, DOI: 10.1016/j.imu.2019.100204.

[5] Deepti Sisodia, Dilip Singh Sisodia, “Prediction of Diabetes using Classification Algorithms”, January 2018, Procedia Computer Science 132:1578-1585, DOI: 10.1016/j.procs.2018.05.122.

[6] Maher Maalouf, "Logistic regression in data analysis: An overview", July 2011, International Journal of Data Analysis Techniques and Strategies 3(3):281-299, DOI: 10.1504/IJDATS.2011.041335. [7] Lishan Wang, "Research and Implementation of Machine Learning Classifier Based on KNN", December 2019, IOP Conference Series Materials Science and Engineering 677:052038, DOI: 10.1088/1757 899X/677/5/052038.

[8] S. Gopal Krishna Patro, Kishore Kumar Sahu, “Normalization: A Pre processing Stage”, March 2015, International Advanced Research Journal in Science, Engineering and Technology, DOI: 10.17148/IARJSET.2015.2305.

[9] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood, “Random Forests and Decision Trees”, September 2012, International Journal of Computer Science Issues 1694-0814, DOI: 10.1.1.402.3863.

[10] Matthew N O Sadiku, Adebowale E. Shadare, Sarhan M Musa, Cajetan Akujuobi, “Data Visualization”, December 2016, International Journal of Engineering Research and Advanced Technology ISSN: 2454-6135.

[11] Pima Indian Diabetes Dataset, accessed on 5th January 2021, UCI Machine Learning depository, https://www.kaggle.com/uciml/pima indians-diabetes-database

[12] Mahmoud, Alamery, Fouad, Altinawi, & Youssef, “ An automatic detection system of diabetic retinopathy using a hybrid inductive machine learning algorithm,” Personal and Ubiquitous Computing, pp 1-15, 2021.

[13] Bernard, D'Elia, Kabadi, & Wong, "An early detection system for foot ulceration in diabetic patients," 2009 IEEE 35th Annual Northeast Bioengineering Conference. IEEE, PP- 1-2, April 2009.

[14] Abràmoff, Michael D., Reinhardt, Russel, Folk, Mahajan, Niemeiier, and Quellec, "Automated early detection of diabetic retinopathy," Ophthalmology, Vol-117,Issue 6,pp- 1147-1154,2010.

[15] Fraiwan, L., AlKhodari, M., Ninan, J., Mustafa, Saleh, & Ghazal, "Diabetic foot ulcer mobile detection system using smart phone thermal camera: a feasibility study," Biomedical engineering online 16.1 (2017): PP 1-19.

[16] Z. R. B, J. Salvador-meneses, and J. Garcia-rodriguez, Preprocessing to Improve Categorical, vol. 1. Springer International Publishing, 2018.

[17] P. Manvitha and N. Swaroopa, "Campus Placement Prediction Using Supervised Machine Learning techniques," Int. J. Appl. Eng. Res., vol. 14, no. 9, pp. 2188–2191, 2019.

[18] M. Nirmala, "Student Academic Performance Prediction under Various Machine Learning Classification Algorithms," Int. J. Res. Appl. Sci. Eng. Technol., vol. 9, no. 11, pp. 221–237, 2021, doi: 10.22214/ijraset.2021.38786.

[19] K. Sumathi, S. Kannan, and K. Nagarajan, "Data Mining: Analysis of student database using Classification Techniques," Int. J. Comput. Appl., vol. 141, no. 8, pp. 22–27, 2016, doi: 10.5120/ijca2016909703.

[20] S. Garcia, J. Luengo, and F. Herrera, Data Preprocessing in Data Mining. Intelligent Systems Reference Library. 2015, vol. 10. 2015.

[21] M. A. H. Farquad and I. Bose, "Preprocessing unbalanced data using support vector machine," Decis. Support Syst., vol. 53, no. 1, pp. 226 233, 2012, doi: 10.1016/j.dss.2012.01.016.

[22] K. Govindasamy and T. Velmurugan, "A Study on Classi fi cation and Clustering Data Mining Algorithms based on Students Academic Performance Prediction," Int. J. Control theory Appl., vol. 10, no. 23, pp. 147–160, 2017.

[23] V. K. Harihar and D. G. Bhalke, "Learning," vol. 12, no. 2, 2020, doi: 10.18090/samriddhi.v12iS2.17.

[24] A. Guezzaz, Y. Asimi, M. Azrour, and A. Asimi, "Mathematical Validation of Proposed Machine Learning Classifier for Heterogeneous Traffic and Anomaly Detection," vol. 4, no. 1, pp. 18 24, 2021, doi: 10.26599/BDMA.2020.9020019.

[25] D. Buenaño-Fernández, D. Gil, and S. Luján-Mora, "Application of machine learning in predicting performance for computer engineering students: A case study ," Sustain., vol. 11, no. 10, pp. 1–18, 2019, learn.org/stable/modules/generated/sklearn.svm.SVR.html doi: 10.3390/su11102833.

[26] E. Pasolli, F. Melgani, and Y. Bazi, "Support Vector Machine Active Learning Through ignificance Space Construction," vol. 8, no. 3, pp. 431–435, 2011.

[27] L. S. Maurya, M. S. Hussain, and S. Singh, "Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on Academic Performance," Appl. Artif. Intell., vol. 35, no. 6, pp. 403–420, 2021, doi: 10.1080/08839514.2021.1901032.

[28] S. C. Gupta, D. K. Singh, and N. Goel, "Enhancing the performance of diabetes prediction using tuning of hyperparameters of classifiers on IMBALANCED," vol. 12, no. 6, pp. 1646–1662.

[29] M. Tajul, R. Ab, and Y. Yusof, "Graduates Employment Classification using Data Mining Approach," vol. 020002, 2016, doi: 10.1063/1.4960842.

[30] Method for Online Learning Scenarios," in 2021 IEEE 4th International Conference on Electronics Technology, ICET 2021, 2021, pp. 1137–1140, doi: 10.1109/ICET51757.2021.9451004.

[31] J. Xu, K. H. Moon, and M. Van Der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," IEEE J. Sel. Top. Signal Process., vol. 11, no. 5, pp. 742 753, 2017, doi: 10.1109/JSTSP.2017.2692560.

[32] B. Perez, C. Castellanos, and D. Correal, "Applying Data Mining Techniques to Predict Student Dropout: A Case Study," 2018 IEEE 1st Colomb. Conf. Appl. Comput. Intell. ColCACI 2018 - Proc., pp. 1–6, 2018, doi: 10.1109/ColCACI.2018.8484847.