Automated Student Result Analysis and Career Guidance System Using Robotic Process Automation and Decision Tree Classification

Sanika Sashte[1], Rajwee Wable[2], Krishna Mare[3]

[1]Sanika Sashte Electronics & Telecommunication AISSMS COE
[2]Rajwee Wable Electronics & Telecommunication AISSMS COE
[3]Krishna Mare Electronics & Telecommunication AISSMS COE

---***---

Abstract - For this project, we created a cutting-edge Robotic Process Automation (RPA) system that would improve and expedite the examination of student performance data. The student result PDFs are effectively converted by the system into an Excel format that is specified, with each row representing a student's results. The system conducts a thorough study of the outcomes using a machine learning technique called the Decision Tree Classification Algorithm. This analysis indicates appropriate career options, like software development or system analysis, for each student. By automating the data entry for eighty pupils, this extensive process takes place concurrently with the PDF-to-Excel conversion, greatly increasing efficiency and saving teachers a great deal of time. Furthermore, the system offers tailored career counseling, supporting educators in counseling students on the best career options and empowering students. By automating the data entry for eighty pupils, this extensive process takes place concurrently with the PDF-to-Excel conversion, greatly increasing efficiency and saving teachers a great deal of time. The system also offers pupils individualized career counseling, which helps teachers counsel students on the best job options and empowers students to make better decisions about their futures. The project provides a solid solution for automated result analysis and career path advice, showcasing the usefulness of RPA and machine learning in educational contexts.

***Key Words*:** Robotic Process Automation ,Analysis, Machine Learning, Excel File, PDF File, Software Bot

1.INTRODUCTION

The combination of technology and education in the era of digital transformation has the power to modify conventional approaches and usher them into a new era of effectiveness and customisation. This fusion is embodied in our project, a groundbreaking Robotic Process Automation (RPA) system that transforms the analysis and interpretation of student results by smoothly combining advanced automation with machine learning.Imagine a system where the laborious process of organizing student outcome PDFs into an Excel format by hand becomes obsolete.

In addition to automating this conversion process, our RPA system uses a Decision Tree Classification Algorithm to do a simultaneous, in-depth study of each student's performance. By analyzing the data, this state-of-the-art machine learning method reveals insights and patterns that could otherwise go unnoticed.There is still more innovation to come. In addition to providing results analysis, our system serves as a guide for students by recommending specific career options, like software development or system analysis, based on their individual academic performance.

With the speed and accuracy with which this individualized career counseling is provided—unprecedented—teachers are better equipped to offer guidance, and students are given the opportunity to carefully explore their options for the future.Suppose there was a situation where teachers could focus on more meaningful teaching exchanges rather of being burdened by the tedious task of entering data for eighty students.

This is made possible by our RPA system, which significantly cuts down on the amount of time spent on administrative work and improves the effectiveness of the learning environment as a whole.This initiative essentially embodies the revolutionary potential of technology in education. We provide a comprehensive solution that enhances the educational experience with data-driven career recommendations while streamlining result processing via the use of RPA and machine learning.

This invention promises a better, more productive future for both students and teachers, demonstrating the limitless opportunities that exist when technology and education come together.

2. Literature Survey

In recent years, there has been a lot of interest in the application of robotic process automation (RPA) and machine learning in educational settings. This review of the literature explores the state of the art in terms of academic process automation and machine learning applications for individualized student counseling.Education and Robotic Process Automation (RPA).

RPA has become a game-changing tool across a number of industries, including education. Willcocks, Lacity, and Craig's (2015) research shows how RPA may automate repetitive, rule-based tasks, improving accuracy and efficiency. RPA has been used at educational institutions to handle administrative duties such as processing results, managing fees, and keeping track of attendance (Aguirre et al., 2017).
These studies demonstrate how RPA can reduce administrative workloads so that teachers can concentrate more on mentorship and instruction.Automated Data Processing. One of the most important areas of research is the transformation of

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

unstructured data—like PDF documents—into structured formats. With the use of optical character recognition (OCR) and natural language processing (NLP), techniques for automatic data extraction from PDFs have substantially evolved (Mederake et al., 2020).

These technologies lay the groundwork for further data analysis by enabling the very accurate extraction of pertinent information.Educational data mining has made extensive use of machine learning methods, namely decision tree classification, to assess student performance and forecast results. In their discussion of machine learning methods for educational data, Romero and Ventura (2010) point out that decision trees are a useful tool for producing intelligible and useful insights.

The subject of personalized career guidance based on academic performance is developing, with studies investigating the application of data-driven methods to suggest appropriate career pathways. According to Santos et al. (2017), algorithms that use student data analysis to recommend jobs have demonstrated promise in connecting students' interests and strengths with possible career paths. These solutions have the combined advantage of raising career preparedness and student happiness
.

3. DESCRIPTION OF ARCHITECTURE

3.1.Decision Tree Algorithm Architecture

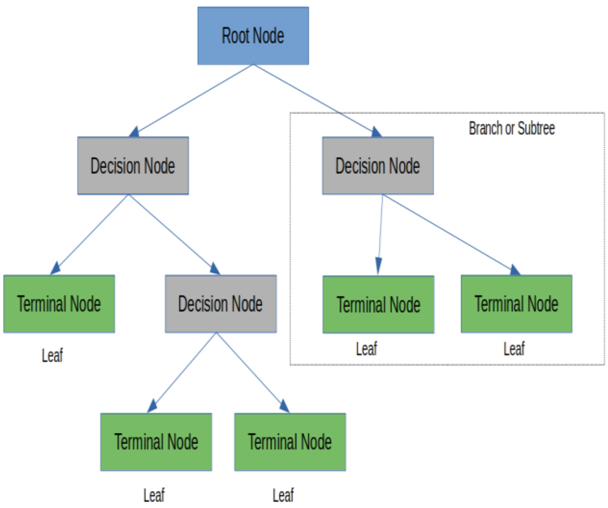


Fig .1

For both classification and regression applications, decision tree classification is a well-liked and simple machine learning method. Because of its ease of use, readability, and capacity to handle both categorical and numerical data, it is very well-liked. An extensive introduction to decision trees is given in this section, along with details on the various kinds of nodes that make up its structure.

Components of a decision tree include the following:

Root Node: In a decision tree, the root node is the highest node and it symbolizes the complete dataset. It serves as the foundation upon which the tree grows.The feature that best divides the data in accordance with a particular criterion (e.g., Gini impurity, Information Gain) is used to divide the dataset at the root node.

Decision Nodes : Decision nodes, sometimes referred to as internal nodes, are locations in the tree where additional data splitting occurs according to particular attributes.A test or condition on an attribute (such as "Is age > 18?") is applied by each decision node, producing two or more branches.Depending on how the test turns out, these nodes send the data in various directions.

Leaf Nodes: The ends of the tree where there is no more splitting are referred to as leaf nodes, terminal nodes, or leaves.In classification trees, each leaf node represents a class label; in regression trees, it represents a value.A categorization rule that can be applied to fresh data points is the route from a root node to a leaf node.

3.2.Splitting Criteria

The process of splitting nodes in a decision tree entails determining which characteristic best separates the data into classes. Common criteria for deciding on the best split include:

Gini impurity is largely used in classification trees (e.g., CART).Determines the impurity or diversity of the data. A lower Gini impurity corresponds to a purer node. Formula: $Gini = 1 - \sum (p_i)^2$ Gini=1−∑(p i) 2 , where p_i p i is the probability of a particular class at a node.

Information Gain:
Based on the concept of entropy in information theory.
Measures the decrease in entropy or uncertainty about the data following a split.
Formula: $InformationGain = Entropy_{parent} - \sum (\frac{N_k}{N} \times Entropy_{child_k})$ InformationGain=Entropy parent −∑(N N k ×Entropy child k), where N_k N k is the number of samples in the child node k k and N N is the total number of samples in the parent node.

Chi-Square:Used to assess the statistical significance of a split.Compares the observed distribution of classes to the expected distribution if there were no attribute-class relationship.

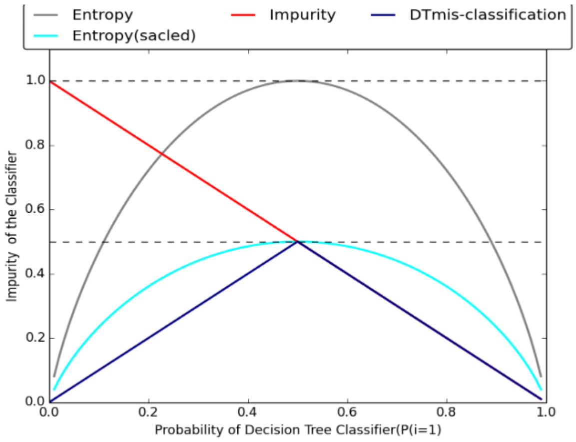


Fig.2

The success rate diagram for a Decision Tree Classification (DTC) algorithm effectively depicts the link between various splitting criteria, such as impurity measurements and entropy, and the probability of successful classification by the decision tree. This section will explain how to present this figure, with the impurity classifier's fields on the y-axis and the decision tree classifier's probability on the x-axis, as well as how entropy is used.

Diagram components include an impurity classifier on the y-axis.

The y-axis depicts the impurity classifier, which determines the cleanliness of the decision tree's nodes. Gini impurity and entropy are two often used impurity measurements. Lower values for these measures suggest purer nodes.

X-Axis: Probability of the Decision Tree Classifier.

The x-axis reflects the decision tree's chance of correctly classifying the data. This is the DTC's success rate, which is commonly expressed as the proportion of correctly identified instances out of all instances.
Entropy

Entropy is one of the factors used to select splits in a decision tree. It measures the level of uncertainty or randomness in data at a node. Lower entropy suggests greater purity and decreased randomness.To get a success rate diagram for the DTC method, perform these steps:

Calculate impurity measures:

Calculate the impurity measure at each node in the decision tree.Plot these numbers on the y axis.Determine the probability of correct classification.

Determine the likelihood of correct classification for the decision tree. This is obtained from the decision tree's accuracy on a validation dataset.Plot these probabilities along the x-axis.

4. Working Principle

4.1 Block Diagram

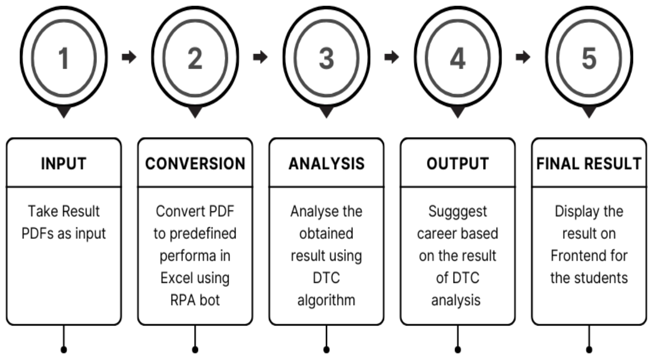


Fig 3

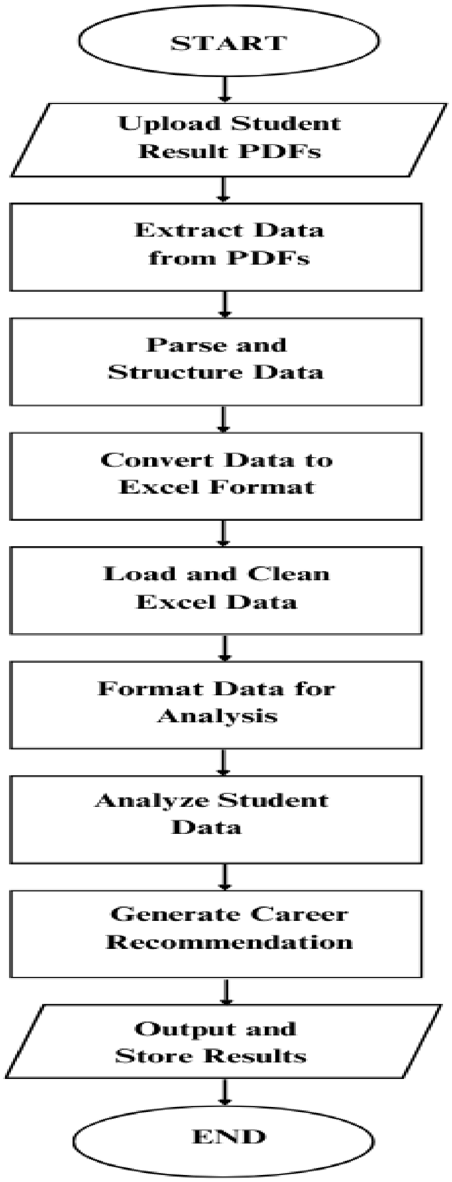
The solution is made up of two primary components: RPA-powered PDF-to-Excel conversion and machine learning-based outcome analysis. The RPA component automates the extraction and conversion of student outcome data from PDFs to a predetermined Excel format. Simultaneously, the ML component uses a Decision Tree Classification Algorithm to analyze the data and recommend appropriate career options for each student.

The RPA system extracts text data from PDFs containing student results using optical character recognition (OCR) technology. It then converts this information into an organized Excel spreadsheet, with each row representing an individual student's findings. The RPA procedure enables a flawless conversion with minimum manual involvement.

After translating the PDFs to Excel, the ML component analyzes the data to find patterns and relationships. The Decision Tree Classification Algorithm considers a variety of characteristics, including academic achievement, topic interests, and extracurricular activities, to suggest appropriate career options for each student. The algorithm's training data consists of past student data and established career paths.
The technology is designed to integrate seamlessly with existing educational procedures. Teachers may start the process by sending PDFs of student results to the system, which will then convert and analyze them automatically. The entire procedure moves quickly, allowing for prompt career advice.

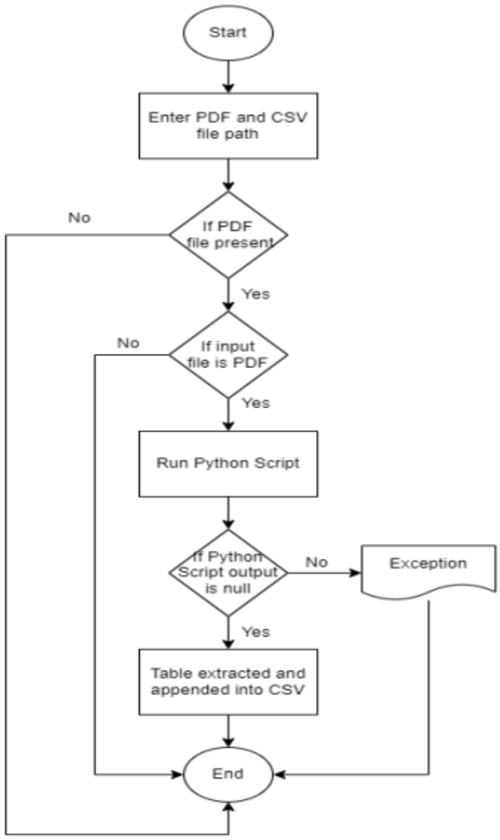
4.2. FLOWCHART

4.2.1 SYSTEM FLOWCHART



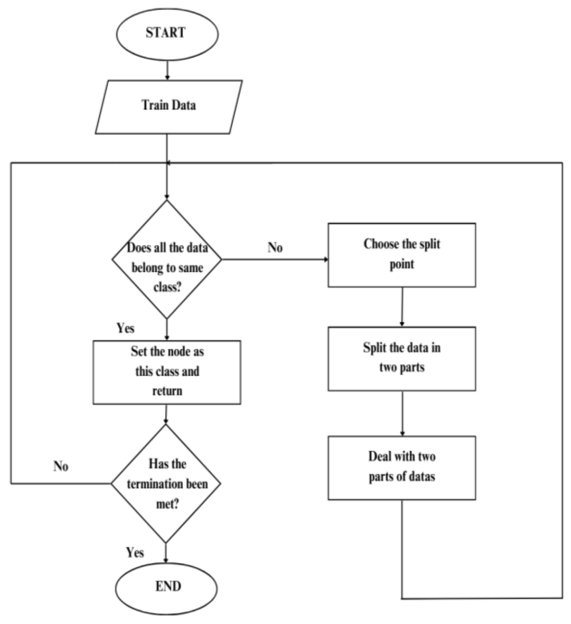
This flowchart represents a comprehensive workflow that starts with uploading student result PDFs and ends with generating and storing personalized career recommendations. The process ensures that data is accurately extracted, cleaned, and analyzed using advanced machine learning techniques, ultimately providing valuable insights to help guide students in their career paths. Each step is crucial for maintaining the integrity of the data and the accuracy of the recommendations.

4.2.2 RPA BOT FLOWCHART



1. Start:
This is the initial point where the process begins.
2. Enter PDF and CSV file path:
The user is prompted to input the file paths for the PDF and CSV files. These paths are necessary for the script to locate and process the files.
3. If PDF file present:
Yes: The flow continues to the next decision point.
No: If no PDF file is present, the process terminates or handles the error (not explicitly shown in the flowchart).
4. If input file is PDF:
Yes: The flow continues to run the Python script.
No: If the input file is not a PDF, the process likely terminates or handles the error (not explicitly shown in the flowchart).
5. Run Python Script:
At this stage, the specified Python script is executed. The script is designed to process the PDF file.
6. If Python Script output is null:
No: If the output is not null (i.e., the script successfully extracted data), the flow continues to the next step.
Yes: If the output is null, an exception is raised.
7. Exception:If the script output is null, an exception is handled. This could involve logging the error, notifying the user, or any other error-handling mechanism defined in your script.
8. Table extracted and appended into CSV:
If the Python script successfully extracts data from the PDF, this data is appended to the CSV file. This means the table data from the PDF is added to the existing data in the CSV file.
9. End:
This signifies the end of the process.

4.2.3 DTC FLOWCHART



This flowchart illustrates the process of building a Decision Tree Classification model, a common machine learning algorithm used for classification tasks. Here's a detailed explanation of each step in the flowchart:
1. Start
Description: The process begins.
2. Train Data
Description: Input the training data into the system.
Purpose: To provide the data that will be used to train the Decision Tree model.
3. Does all the data belong to the same class?
Description: Check if all the data points in the current node belong to the same class.
Yes Branch:
Set the node as this class and return:
If all data points belong to the same class, label this node with that class.
Purpose: To finalize this node as a leaf node with a class label.
No Branch:
Choose the split point:
Identify the best attribute to split the data, which maximizes the separation of different classes.
Purpose:To determine the most informative feature for dividing the data.
Split the data into two parts:
Divide the dataset into subsets based on the chosen split point.
Purpose: To create branches that represent the outcomes of the split.
4. Deal with two parts of data
Description: Recursively apply the decision tree algorithm to the two subsets of data created from the split.
Purpose: To continue building the tree by handling each subset independently, ensuring that the decision tree captures the structure of the data.
5. Has the termination been met?
Description: Check if the termination condition for the algorithm has been met. Termination conditions can include:
A maximum tree depth.
A minimum number of data points in a node.
If further splitting doesn't improve classification significantly.
Yes Branch:
End: Stop the algorithm as the termination condition has been met.
Purpose: To finalize the tree-building process.
No Branch:
Continue the process of checking if all data belongs to the same class or choosing new split points and dealing with subsets.
The flowchart depicts a recursive process of building a Decision Tree Classification model.
5. Termination Check: After processing splits, check if a termination condition is met. If it is, end the process; otherwise, continue.
This process ensures that the decision tree captures the structure and patterns in the data, allowing it to make accurate classifications.

5. APPLICATIONS

Performance Tracking and Reporting:
The system can continuously track student performance over time, generating periodic reports that highlight improvements or declines in academic achievement. These reports can be used by educators to adjust teaching methods and provide targeted support to students who need it.

Student Placement Assistance:
For higher education institutions, the system can aid in student placement processes by matching students with suitable internships and job opportunities based on their academic performance and career interests. This can enhance employability and career readiness.

Curriculum Development:
Insights derived from the detailed analysis of student performance can inform curriculum development. Educators can identify which parts of the curriculum are most effective and which need revision, ensuring that the educational content remains relevant and engaging.

Early Warning Systems:
The system can be configured to detect early signs of academic struggles or disengagement among students. By identifying at-risk students early on, educators can intervene with appropriate measures such as counseling, tutoring, or mentoring programs.

Resource Allocation:
The analysis provided by the system can guide administrators in the allocation of resources such as teaching materials, technological tools, and support staff. This ensures that resources are directed where they are most needed to support student success.
Accreditation and Compliance:
The system can assist educational institutions in maintaining accreditation standards and compliance with educational regulations by providing accurate and up-to-date performance data, which can be easily compiled and submitted for reviews.

6. FUTURE SCOPE

1. Enhanced Machine Learning Models:
Incorporating Advanced Algorithms: Future iterations of the system could integrate more sophisticated machine learning models, such as neural networks or ensemble methods, to improve the accuracy and depth of analysis.
Continuous Learning: Implementing continuous learning algorithms that update and improve the model based on new data can help the system adapt to changing trends in student performance.

2. Integration with Educational Platforms:
LMS Integration:The system could be integrated with Learning Management Systems (LMS) like Moodle or Blackboard, allowing for seamless data transfer and broader analysis of student engagement and performance.
Real-Time Feedback: Developing features for real-time feedback and analysis could help educators intervene more promptly to support students who are struggling.

3. Expanded Career Guidance:
Comprehensive Career Pathways: The system could be expanded to provide more detailed career pathways, including information on necessary skills, required education, and potential job opportunities.
Soft Skills and Extracurricular Analysis: Including analysis of soft skills and extracurricular activities in career recommendations can provide a more holistic view of a student's potential career paths.

4. Personalized Learning Plans:
Adaptive Learning: By analyzing performance data, the system could recommend personalized learning plans tailored to each student's strengths and weaknesses, promoting individualized learning experiences.
Progress Tracking: Implementing features that track students' progress over time and adjust learning plans accordingly can help optimize educational outcomes.

5. Cross-Institutional Data Analysis:
Benchmarking: The system could facilitate benchmarking by comparing data across multiple institutions, providing insights into broader educational trends and standards.
Collaborative Research: Enabling institutions to share anonymized data for collaborative research can drive advancements in educational methodologies and policies.

7. CONCLUSIONS

Finally, the combination of Robotic Process Automation (RPA) and Machine Learning (ML) approaches has resulted in the creation of an efficient and personalized system for analyzing student results and providing career assistance. By automating the translation of student result PDFs into structured Excel forms while also doing in-depth analyses using the Decision Tree Classification Algorithm, the system simplifies operations and delivers significant insights for both educators and students. The project's effective deployment demonstrates its potential to transform the way educational institutions manage student data and provide targeted career advising. Moving ahead, additional refining and extension of the system may result in even greater efficiency and improved educational outcomes.

ACKNOWLEDGEMENT

Prof. N.P. Mawale, our respected Project Guide, provided important direction and assistance to ensure the project's success. His experience, support, and constant devotion were important in designing our project.We would also want to express our profound appreciation to Prof. R.R. Itkarkar, our committed Project Coordinator, who provided continual supervision, insightful criticism, and rigorous attention to detail throughout the project.Their leadership and support have been important in our path, and we are very grateful for their efforts.

REFERENCES

1. Willcocks, L., Lacity, M., & Craig, A. (2015). Robotic Process Automation: The Next Transformation Lever for Shared Services. Journal of Information Technology Teaching Cases, 5(2), 105-113.

2. Aguirre, S., Rodriguez, A., & Vasquez, J. (2017). Robotics Process Automation: A Literature Review. International Conference on Software Process Improvement, 68-74.

3. Mederake, L., Groth, M., & Meinel, C. (2020).A Survey on Optical Character Recognition Systems. Proceedings of the 2020 International Conference on Artificial Intelligence, 75-82.

4. Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.

5. Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2004). Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. Applied Artificial Intelligence, 18(5), 411-426.

6. Santos, O. C., Boticario, J. G., & Perez-Marin, D. (2017).Extending Decision Tree Classification Algorithms for the Use in Intelligent Tutoring Systems. Expert Systems with Applications, 90, 303-314.

7. van der Aalst, W. M. P., Bichler, M., & Heinzl, A. (2018). Robotic Process Automation. Business & Information Systems Engineering, 60(4), 269-272.

8. Heredia, J. A., Vargas, M. F., & Castro, J. J. (2018). An Automated Grading System for Virtual Learning Environments Based on Machine Learning. Computers & Education, 124, 31-42.

9. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and Regression Trees. Wadsworth International Group

10. Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning, 1(1), 81-106.

11. Han, J., Pei, J., & Kamber, M. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.

12. Mohapatra, S., & Bhardwaj, A. (2019). Application of Robotic Process Automation in Industry. Journal of Advanced Research in Dynamical and Control Systems, 11(5), 68-73.

13. Narayana, K. V., & Ravi, V. (2017). Decision Tree Model for Predicting Student Performance. International Journal of Applied Engineering Research, 12(24), 15344-15350.

14. Alwidian, J., & Khasawneh, M. T. (2018). Enhancing Student Performance Prediction Using Decision Tree Classifier and Support Vector Machine. Journal of Theoretical and Applied Information Technology.96(6), 1638-1649.

15. Lacity, M. C., & Willcocks, L. P. (2016).A New Approach to Automating Services. *MIT Sloan Management Review*, 58(1), 41-49.

BIOGRAPHIES (Optional not mandatory)

Prof. N.P. Mawale is a distinguished academic and project guide in the Electronics and Telecommunication Engineering Department at All India Shri Shivaji Memorial Society's College of Engineering (AISSMS COE). With a strong background in electronics and telecommunication, Prof. Mawale has contributed significantly to the field through both his teaching and research endeavors.

Miss Sanika Sashte , is a final year engineering student in electronics and telecommunication engineering at AISSMS COE , having expertise in Java Backend Development.

Miss Rajwee Wable , is a final year engineering student in electronics and telecommunication engineering at AISSMS COE , having expertise in Automation Development.

Mr Krishna Mare , is a final year engineering student in electronics and telecommunication engineering at AISSMS COE , having expertise in Data Analysis.