

Capstone Project Documentation:

Name : A.SURESH KUMAR

College : Institute of data

Date : 19.10.2024



Contents:

1. The project title
2. The project purpose
3. The project objectives and goals
4. Stakeholders' identification
5. Project risks
6. Project benefits
7. Data analysis and modelling
8. The solution for project problems.

1. Project Title:

How can educational institutions leverage predictive analytics to identify at-risk students early in their academic journey, thereby reducing dropout rates and enhancing overall student success.

2.The project purpose:

This project aims to investigate and apply predictive analytics in educational settings in order to detect students who are at risk at an early stage of their academic careers.

The project's goal is to provide a methodical strategy to identifying probable dropout warning signals by analyzing previous data on student performance, attendance, and involvement.

The project aims to improve overall student success and lower dropout rates through focused interventions and individualized support plans.

In the end, this project seeks to establish a more welcoming and encouraging learning environment that encourages student achievement and retention, ultimately enhancing the long-term prosperity of both the institution and its constituents.

3. Project objectives and goals:

- Students:

The primary beneficiaries of the project; their engagement and feedback are critical.

- Parents/Guardians:

Interested in their children's academic success and well-being.

- Regulatory Bodies:

Organizations that may require compliance with educational standards and reporting.

- Community Organizations:

Local groups that may support or be impacted by student success initiatives.

4. Identification Process:

1. Review Project Documentation:

Start with the project charter and existing documentation to compile a list of potential stakeholders.

2. Conduct Interviews:

Engage with key individuals within the institution to gather insights on who else should be considered stakeholders.

3. Utilize Surveys:

Distribute surveys to collect broader input from faculty, staff, and students regarding stakeholder perspectives.

4. Create a Stakeholder Register:

Document each stakeholder's role, interests, influence, and expectations to facilitate ongoing engagement.

Importance of Stakeholder Engagement:

Stakeholder engagement and identification must be done effectively for:

- Being aware of various viewpoints regarding student achievement.
- Making certain that interventions meet the needs of stakeholders.
- Increasing the project's support at all institutional levels.
- The project may increase its efficacy in utilizing predictive analytics to assist at-risk kids by methodically identifying and involving stakeholders, which will ultimately improve student outcomes and retention rates.

5. Project Risks:

There are several hazards associated with using predictive analytics in educational settings to identify kids who are at risk, and these risks must be properly addressed. Important project hazards consist of:

1. Data Privacy and Security Concerns:

There are serious privacy and security concerns when sensitive student data is used. Unauthorized access to personal data could result from improper security, which could have negative legal effects and erode parental and student trust.

2. Model Accuracy and Bias:

The quality of the data used to train predictive models determines how well they perform. Predictions that are erroneous due to incomplete or biased data may fail to identify pupils who are at danger or fail to recognize those who require assistance. This may strengthen preexisting prejudices and result in unsuccessful treatments.

3. User Perception and Trust Issues:

Faculty and students may have doubts about predictive analytics, especially when it comes to the exploitation of their data. When predictive methods lack transparency, trust may be damaged and opposition to the use of such systems may result.

4. Resource Allocation and Management Challenges:

Predictive analytics implementation calls for a large investment in technical infrastructure and qualified staff. Project delays or failure may result from improper management or allocation of these resources.

5. Dependence on Data Quality:

The quality and accessibility of data have a major impact on predictive analytics' efficacy. Inadequate data management techniques can jeopardize the analysis's integrity and produce untrustworthy results.

Educational institutions can increase their chances of successfully utilizing predictive analytics to effectively help at-risk students by proactively identifying and addressing these hazards.

6. Project Benefits:

Predictive analytics implementation in educational institutions has several advantages that can greatly improve institutional efficacy and student achievement. Important advantages include:

1. Early Identification of At-Risk Students:

Early on in their academic careers, schools can use predictive analytics to identify children who could be at danger of dropping out or performing poorly. Timely interventions are made possible by this proactive approach, which raises retention rates and enhances total student achievement.

2. Targeted Support and Interventions:

Institutions can offer individualized support that is catered to the unique requirements of at-risk students by analysing a variety of data points, including attendance, grades, and engagement. Students are given the support they require to maintain their academic progress thanks to this focused guidance.

3. Enhanced Resource Allocation:

Institutions can better allocate resources, like as financial aid and academic support services, by using predictive analytics. Schools can make data-driven decisions that increase operational effectiveness and improve student outcomes by comprehending trends and patterns in student behaviour.

4. Improved Enrolment Management:

Predictive analytics can help institutions anticipate enrolment trends and efficiently manage their recruitment efforts. To satisfy institutional goals and guarantee a broad and well-balanced student body, this capacity enables schools to match their offers with student demand.

5. Continuous Improvement of Educational Quality:

Institutions can pinpoint areas where their course offerings and instructional strategies need to be improved by using predictive models to track student performance and engagement. A culture of constant improvement in the quality of education is fostered by this continuing evaluation.

6. Data-Driven Decision Making:

At every level of the organization, strategic decisions are supported by tangible data from predictive analytics. By giving stakeholders, the capacity to gauge the results of programs and interventions, this data-driven approach promotes effective change management and increases responsibility.

Educational institutions can foster a more encouraging environment for students by utilizing these advantages, which will eventually increase retention rates, boost academic achievement, and increase overall institutional efficacy.

7.Data Analysis and Modelling:

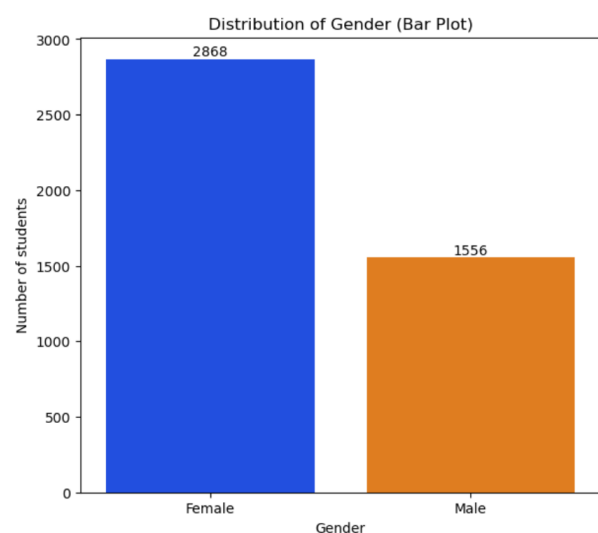
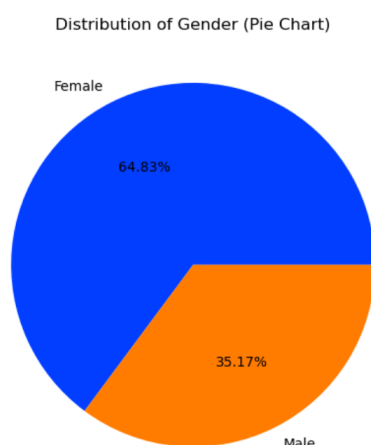
Gender Enrolment Statistics

- **Total Students:** 4,424
- **Female Enrolment:** 2,868 (approximately 65%)
- **Male Enrolment:** 1,556 (approximately 35%)

Key Insights

- **Female Dominance:** The enrolment of females significantly exceeds that of males, indicating a trend towards higher female participation in this educational setting.
- **Implications for Support Programs:** Understanding the gender distribution can help tailor academic support and retention programs to address the specific needs of each group.

These statistics highlight the importance of considering gender dynamics in educational research and policy-making.



Course Outcome Statistics:

- **Total Students:** 4,424
- **Graduates:** 3,003 (approximately 68%)
- **Dropouts:** 1,421 (approximately 32%)

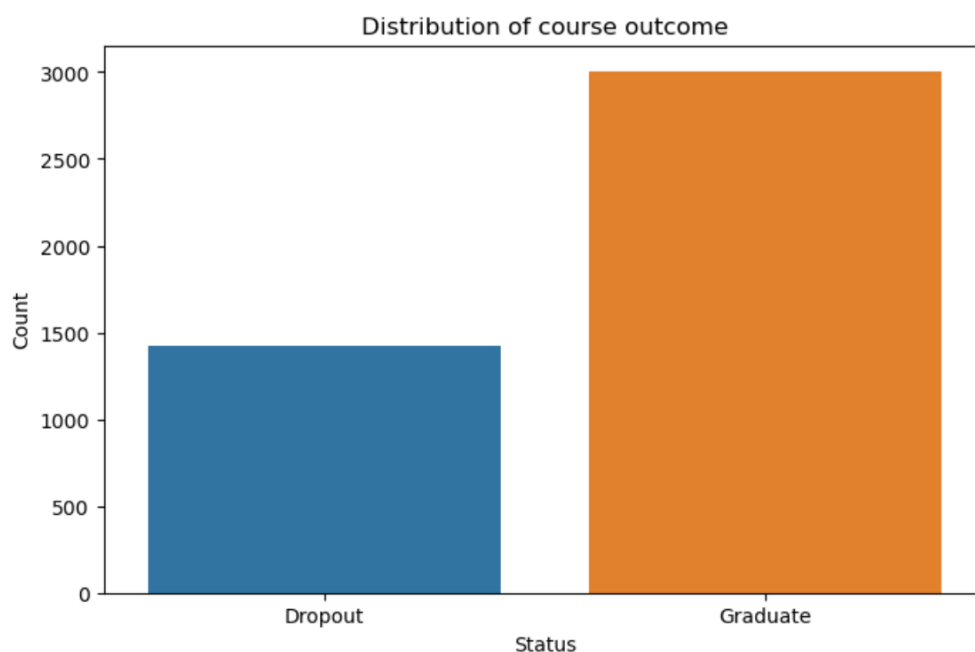
Key Insights

- **High Graduation Rate:** With 68% of students graduating, the institution demonstrates a relatively strong retention and success rate.
- **Dropout Rate:** The dropout rate of 32% indicates a significant number of students who do not complete their courses, highlighting areas for potential intervention.

Implications

- **Support Strategies:** The data suggests a need for targeted support programs to reduce dropout rates and enhance student retention.
- **Further Analysis:** Understanding the factors contributing to dropout can help improve academic outcomes and develop proactive measures for at-risk students.

-



Scholarship Receiver Statistics:

- **Total Students:** 4,424
- **Students Receiving Scholarships:** 1,099 (approximately 25%)
- **Students Not Receiving Scholarships:** 3,325 (approximately 75%)

Key Insights

- **Low Scholarship Uptake:** Only 25% of students receive scholarships, suggesting that a significant majority are either not applying for or not qualifying for financial aid.
- **Potential Barriers:** The high number of students not receiving scholarships may indicate barriers related to awareness, application processes, or eligibility criteria.

Implications

- **Enhancing Access:** Institutions could benefit from initiatives aimed at increasing scholarship awareness and accessibility to support more students financially.
- **Impact on Outcomes:** Understanding the relationship between scholarship availability and student success may provide insights into improving graduation rates and reducing dropout rates.

These statistics underscore the importance of financial support in higher education and the need for effective outreach strategies.

Scholarship Recipients and Non-Dropout Rates:

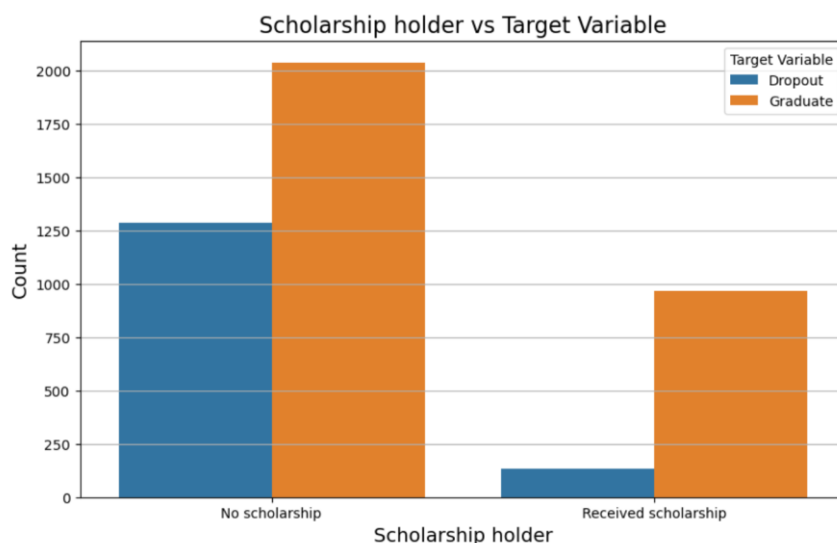
Scholarship recipients generally exhibit higher retention rates compared to non-recipients, indicating a strong correlation between financial support and academic persistence. Here are some key points regarding this relationship:

- **Higher Retention Rates:**

Studies show that scholarship recipients often have retention rates significantly above those of non-recipients. For example, one study reported a retention rate of **90.6%** for scholarship recipients compared to **84.1%** for non-recipients.

- **Impact of Financial Support:**

Scholarships can alleviate financial burdens, allowing students to focus more on their studies rather than on financial stressors. This support is crucial for underfunded students, as evidenced by programs that have improved retention outcomes through targeted financial aid.



Tuition Fees and Student Outcomes Key Statistics:

- **Tuition Fees Not Up to Date (NO):**

- Dropouts: 457
- Graduates: 71
- Total: 528

- **Tuition Fees Up to Date (YES):**

- Dropouts: 964
- Graduates: 2932
- Total: 3896

Analysis:

1. **Correlation with Academic Success:**

- Students with up-to-date tuition fees have a significantly higher graduation rate (75.3%) compared to those who are behind on payments (13.4%).
- This suggests a strong positive correlation between timely fee payment and academic success.

2. **Dropout Risk:**

- The dropout rate for students not up to date with fees is alarmingly high at 86.6%.
- In contrast, only 24.7% of students with up-to-date fees drop out.

3. **Financial Stability Impact:**

- The data implies that financial stability, as indicated by timely fee payments, plays a crucial role in a student's ability to complete their education.

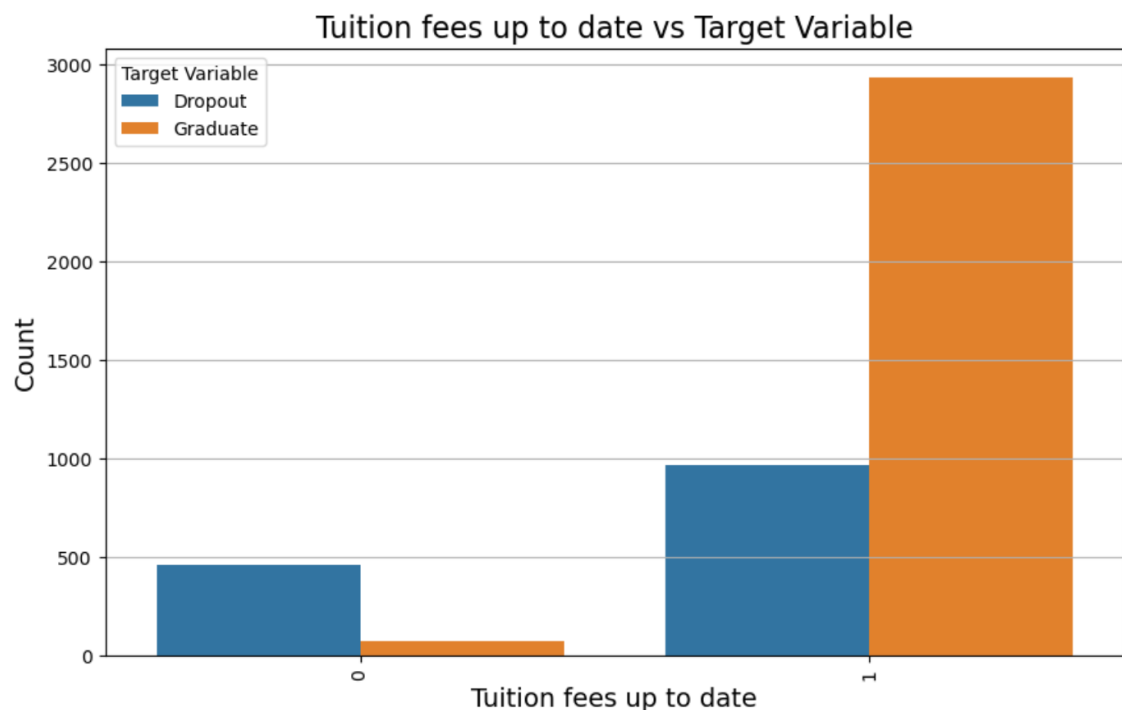
4. Potential Interventions:

- These findings highlight the need for targeted financial support and early intervention for students struggling with tuition payments.
- Implementing flexible payment plans or additional financial aid could significantly improve retention rates.

5. Overall Trends:

- The majority of students (88%) are up to date with their tuition fees, which correlates with the overall higher graduation numbers.

This data underscores the critical link between financial stability and academic success, suggesting that addressing financial challenges could be a key strategy in improving graduation rates and reducing dropouts.



Model comparison:

Model Performance Summary:

1. Random Forest

- Best Score: 0.924853 (92.49%)
- Best Parameters: max_depth=20, n_estimators=200
- Highest performing model

2. Support Vector Machine (SVM)

- Best Score: 0.882804 (88.28%)
- Best Parameters: C=10, gamma=0.1, kernel='rbf'
- Second-best performer

3. Decision Tree

- Best Score: 0.875727 (87.57%)
- Best Parameters: criterion='entropy'
- Third in performance

4. Logistic Regression

- Best Score: 0.853660 (85.37%)
- Best Parameters: C=5
- Fourth in ranking

5. Naive Bayes (Gaussian)

- Best Score: 0.794127 (79.41%)
- No specific parameters optimized
- Lowest performing among the compared models

Key Insights:

1. **Top Performer:** Random Forest outperforms other models significantly, suggesting it is best suited for this dataset.
2. **Complexity vs. Performance:** More complex models (Random Forest, SVM) generally performed better than simpler ones (Naive Bayes, Logistic Regression).
3. **Hyperparameter Tuning:** Most models benefited from hyperparameter optimization, highlighting the importance of this step-in model development.
4. **Performance Range:** There is a notable performance gap (about 13%) between the best and worst models, indicating the choice of model is crucial for this dataset.
5. **Potential for Ensemble:** Given the varied performances, an ensemble approach combining these models might yield even better results.

This comparison provides valuable insights for selecting the most appropriate model for predicting student dropout and academic success, with Random Forest emerging as the standout choice for this dataset.

Feature selection:

To address the business question, I've used these below features so that drop out can be predicted very early in enrolment process.

X = 'Admission grade', 'Age at enrollment', 'Previous qualification (grade)', 'Tuition fees up to date',
'Course', 'GDP', 'Mothers_Occupation',
'Fathers_Occupation', 'Mothers_Qualifications',
'Fathers_Qualifications', 'Target'

Y='Target'

Confusion matrix Interpretation:

Breakdown of Values:

1. **True Negatives (TN):** 516
 - The model correctly predicted 516 instances as negative.
2. **False Positives (FP):** 70
 - The model incorrectly predicted 70 instances as positive when they were negative.
3. **False Negatives (FN):** 82
 - The model incorrectly predicted 82 instances as negative when, they were positive.
4. **True Positives (TP):** 534
 - The model correctly predicted 534 instances as positive.

Summary of Metrics:

- **Accuracy:** $\approx 87.4\%$
- **Precision:** $\approx 88.4\%$
- **Recall:** $\approx 86.6\%$
- **F1 Score:** $\approx 87.3\%$

Insights:

- **High Overall Performance:** The model demonstrates strong performance across all metrics, with over 87% in accuracy, precision, recall, and specificity.
- **Balanced Predictions:** The model shows similar performance in predicting both classes, as evidenced by the balanced number of true positives and true negatives.
- **Low Error Rates:** Both false positives (70) and false negatives (82) are relatively low, indicating good discrimination ability.
- **Potential for Real-World Application:** Given its high accuracy and balanced performance, this model could be valuable for predicting student outcomes and implementing targeted interventions.

This confusion matrix suggests a highly effective classification model that could be reliably used in educational settings for predicting student success or identifying at-risk students.

8. The Solutions for Project Problems:

To effectively address the potential problems associated with implementing predictive analytics in educational institutions, the following solutions can be adopted:

1. Data Privacy and Security Measures

- **Implement Robust Security Protocols:** Utilize encryption, secure access controls, and regular security audits to protect sensitive student data from unauthorized access.
- **Establish Clear Data Governance Policies:** Create policies that outline data usage, storage, and sharing practices to ensure compliance with legal regulations such as FERPA (Family Educational Rights and Privacy Act) and GDPR (General Data Protection Regulation).

2. Ensuring Model Accuracy and Reducing Bias

- **Regularly Review and Update Models:** Continuously refine predictive models using diverse and representative datasets to minimize bias and enhance accuracy.
- **Incorporate Feedback Mechanisms:** Establish channels for educators and students to provide feedback on model predictions, allowing for adjustments based on real-world outcomes.

3. Building Trust and Transparency

- **Enhance Communication Strategies:** Clearly communicate the purpose, benefits, and processes of predictive

analytics to students, faculty, and parents to build trust and alleviate concerns.

- **Involve Stakeholders in the Process:** Engage students and faculty in discussions about data usage and decision-making processes to foster a sense of ownership and transparency.

4. Effective Resource Management

- **Conduct a Needs Assessment:** Evaluate the necessary resources (e.g., personnel, technology) before project implementation to ensure adequate support is available.
- **Utilize Pilot Programs:** Start with small-scale pilot programs to test predictive analytics tools before full-scale implementation, allowing for adjustments based on initial findings.

5. Enhancing Data Quality

- **Establish Data Management Practices:** Implement standardized data collection methods and regular audits to maintain high data quality.
- **Train Staff on Data Entry Procedures:** Provide training for faculty and administrative staff on accurate data entry practices to minimize errors.

6. Fostering Continuous Improvement

- **Create an Evaluation Framework:** Develop metrics to assess the effectiveness of interventions based on predictive analytics insights. Regularly review these metrics to identify areas for improvement.

- **Encourage a Culture of Feedback:** Promote an environment where continuous feedback is valued, allowing educators to adapt teaching methods based on student performance data.

By proactively implementing these solutions, educational institutions can mitigate potential challenges associated with predictive analytics, ensuring a smoother implementation process that ultimately enhances student success and institutional effectiveness.