**Standard Operating Procedure (SOP) for UConn Capstone Project**

**Title: Tracking Student Journey: Analyzing Retention and Graduation Rates**

**1. Purpose**

The **Tracking Student Journey** project was undertaken as part of the UConn Capstone Program to analyze and improve student retention and graduation rates. By leveraging data-driven methodologies, the project aimed to uncover key factors influencing student success and provide strategic recommendations to enhance retention outcomes.

This **Standard Operating Procedure (SOP)** serves as a guide to document the processes followed, ensuring clarity, consistency, and standardization in project execution. It details the methodologies applied for **data collection, preprocessing, modeling, visualization, and business insights**, while also outlining the roles and responsibilities of the team members.

**2. Project Team and Responsibilities**

The success of this project was driven by a dedicated team of professionals, each bringing unique expertise to different aspects of the study. Below is an overview of the **UConn Capstone Project Team** and their specific roles:

**Project Leadership:**

* **Bajram Lumani (Project Manager)**
  + Led the overall project execution, ensuring alignment with business objectives.
  + Managed stakeholder communications and facilitated team collaboration.
  + Provided guidance on strategic decision-making and resource allocation.
* **Shriya S M (Sponsor Liaison)**
  + Acted as the bridge between stakeholders and the project team.
  + Gathered sponsor requirements and ensured project deliverables met expectations.
  + Provided insights into institutional priorities and strategic goals.

**Technical Team:**

* **B V S L Saikrishna (Technical Lead)**
  + Managed data architecture and infrastructure.
  + Provided technical guidance on data integration, storage, and preprocessing.
  + Ensured seamless data pipeline execution and model implementation.
* **Rohit Reddy M (Data Analyst)**
  + Performed data extraction, cleaning, and transformation.
  + Developed key insights through statistical analysis and visualizations.
  + Created interactive **Power BI dashboards** for real-time decision-making.
* **Siva Sobhan Prabhala (Data Scientist)**
  + Developed and fine-tuned machine learning models to predict student retention.
  + Evaluated model performance and optimized predictive accuracy.
  + Assisted in feature selection and engineering to enhance predictive capability.
* **Deepak Pasala (Business Analyst)**
  + Conducted gap analysis and business process mapping.
  + Translated data-driven insights into actionable recommendations for stakeholders.
  + Assisted in the documentation of findings and strategic proposals.

Each team member played a vital role in ensuring the project’s success, utilizing their specialized skills to drive impactful analysis and meaningful insights.

**3. Scope**

This SOP applies to all tasks associated with the **Tracking Student Journey Project**, including:

* **Data collection & integration** from multiple sources.
* **Data preprocessing & cleaning** to ensure quality and consistency.
* **Exploratory data analysis (EDA)** to uncover patterns and trends.
* **Machine learning modeling** to predict retention and graduation rates.
* **Data visualization & reporting** through Power BI dashboards.
* **Business recommendations** to enhance student retention strategies

**4. Data Collection and Integration**

**4.1 Data Sources**

The project utilized multiple datasets to gain comprehensive insights:

* **Student Retention Data** – Historical student retention and graduation records.
* **Student GPA Tracking Data** – Academic performance metrics.
* **Class-Taking Data** – Course enrollment trends.
* **Sixty-Second Survey Data** – Student engagement and satisfaction levels.
* **Q\_Center Sign-ins Data** – Academic support and tutoring attendance.

**4.2 Data Merging Strategy**

To create a unified dataset, **inner joins** were performed between multiple datasets, ensuring only relevant records were retained.

* **Pivoting & Aggregation** – Addressed redundancy by restructuring data based on student IDs.
* **GroupBy Functions** – Summarized features to remove unnecessary duplication.

**5. Data Preprocessing and Cleaning**

**5.1 Handling Missing Values**

* Removed columns with **>30% missing values**.
* Used **mean imputation** for numerical data gaps.
* Assigned categorical missing values to the most common occurrence.

**5.2 Feature Engineering**

* Created new features based on student engagement and academic patterns.
* Standardized numerical features using **Min-Max Scaling** for better model performance.

**5.3 Data Splitting**

* **80% Training Data**
* **20% Testing Data**

**6. Exploratory Data Analysis (EDA)**

**6.1 Key Findings**

* **Pell Grants** significantly impacted student retention.
* **Student Organization Participation** showed a positive correlation with retention.
* **Academic Confidence & Stress Levels** influenced graduation outcomes.

**6.2 Visualizations**

* **Bar Charts** – Retention trends based on financial aid.
* **Heatmaps** – Correlation between stress, GPA, and retention.
* **Boxplots** – Distribution of retention scores across departments.

**7. Machine Learning Models**

**7.1 Model Selection**

The following models were trained and evaluated:

* **Logistic Regression**
* **Decision Trees**
* **Random Forest**
* **Gradient Boosting (XGBoost)**

**7.2 Model Evaluation**

* **Accuracy, Precision, Recall, and F1-Score** were used for performance measurement.
* **Gradient Boosting** emerged as the best-performing model with **highest accuracy**.

**8. Business Insights and Recommendations**

**8.1 Financial Support Strategies**

* Introduce **on-campus part-time employment** to ease financial burdens.
* Provide **financial counseling** to Pell Grant recipients to enhance retention.

**8.2 Enhancing Student Engagement**

* Allocate **increased funding** to student organizations.
* Recognize and **reward club participation** to foster engagement.

**8.3 Academic Program Initiatives**

* Promote UConn’s **unique curriculum and career placement stats** to improve retention.
* Implement **monthly alumni mentorship programs** to guide first-year students.

**9. Visualization and Reporting**

* **Power BI dashboards** were created for interactive analysis and tracking.
* Final reports were compiled for stakeholders to support strategic decision-making.

**10. Tools and Technologies Used**

* **Data Processing**: Python (Pandas, NumPy, Seaborn, Scikit-learn)
* **Machine Learning Models**: XGBoost, Random Forest, Logistic Regression
* **Data Visualization**: Power BI, Matplotlib, Seaborn
* **Project Management**: Jira, MS Project

**11. Conclusion**

The **Tracking Student Journey Project** successfully identified **key factors influencing student retention** and provided actionable recommendations to **enhance student success at UConn**.

This SOP serves as a structured guideline for future projects in higher education analytics, ensuring that **data-driven methodologies** continue to drive institutional improvements.

📌 **Final Deliverables:**  
✅ Power BI Dashboard  
✅ Machine Learning Model Code (Python)  
✅ Business Recommendations Report

This document will be maintained for **future reference and continuous improvement** in retention analysis.