**JPMC Data Science Bootcamp Capstone Project**

**Objective:**

The objective of this capstone project is to provide students an opportunity to showcase their ability to work on a real-world data science problem from data preprocessing to model deployment. Students are required to select their own dataset, process it, develop a neural network model, and create a user interface (UI) to demo the final product.

**Requirements:**

**1. Dataset Selection:**

* Students must select their own dataset.
* The dataset should have a minimum of 50,000 records and at least 20 features.
* The dataset should preferably be related to a business or social problem, ensuring the project has business or social value.

**2. Data Preprocessing:**

a. Missing Values:

* Handle missing values using appropriate techniques (e.g., imputation, deletion).
* Document the percentage of missing values for each column and the technique used to handle them.

b. Outliers:

* Identify and treat outliers in the dataset.
* Use visualizations like boxplots to showcase the identification of outliers.

c. Resampling:

* If the dataset is imbalanced, use resampling techniques such as oversampling, undersampling, or using the Synthetic Minority Over-sampling Technique (SMOTE).

d. Feature Selection and Importance:

* Use appropriate techniques to select relevant features (e.g., Recursive Feature Elimination).
* Determine the importance of features using techniques like Permutation Importance or SHAP values.

e. Creating New Columns/Merging Datasets:

* If relevant, engineer new features that can aid in model performance.
* Merge additional datasets if it provides added value to the project.

**3. Model Development:**

* Use neural networks for modeling.
* Implement and tune the neural network architecture.
* Use appropriate evaluation metrics to judge the performance of the model.
* Document the architecture and hyperparameters used.

**4. User Interface (UI) Development:**

* Develop a simple UI where users can input data and get predictions from the trained model.
* The UI should be intuitive and user-friendly.
* Bonus: Include visualizations on the UI to display the model's predictions or other relevant insights.

**5. Presentation (12 minutes):**

a. Business Value:

* Begin by explaining the business or social value of the chosen problem.
* Discuss potential impact and benefits of solving the problem.

b. Process of Developing the Model:

* Walk through the data preprocessing steps.
* Discuss challenges faced and how they were addressed.

c. Tech Choices:

* Discuss the technology stack used, including tools, libraries, and frameworks.
* Explain the rationale behind choosing a neural network for modeling.

d. Model Performance:

* Present the performance metrics of the neural network.
* Discuss any overfitting/underfitting issues and how they were addressed.

e. Demonstration:

* Conclude with a short demonstration of the developed UI.
* Show how the model works in real-time and how users can benefit from it.

**Deliverables:**

1. **Codebase**: Clean and well-documented code for data preprocessing, modeling, and UI development.
2. **Report**: A detailed report explaining all the steps, decisions, and results.
3. **Presentation Slides**: A slide deck for the 12-minute presentation.
4. **Demo**: A working demonstration of the model through the developed UI.

**Evaluation Criteria:**

1. **Relevance of Dataset**: Importance of the business/social problem tackled.
2. **Data Preprocessing**: Thoroughness and appropriateness of preprocessing techniques.
3. **Model Architecture & Performance**: Complexity, appropriateness, and performance of the neural network.
4. **UI**: Usability and functionality of the user interface.
5. **Presentation**: Clarity, structure, and effectiveness of the presentation.

**Recommendations:**

1. **Dataset Selection**: Datasets from platforms like Kaggle, UCI Machine Learning Repository, or government databases can be good starting points.
2. **Technology Stack**: Python with libraries like Pandas, Scikit-learn, TensorFlow/Keras for data preprocessing and modeling. For UI, frameworks like Flask or Streamlit can be useful.
3. **Collaboration**: While individual efforts are appreciated, students can also consider teaming up to distribute tasks and bring diverse skill sets to the project.

**Conclusion:**

This capstone project is designed to test the comprehensive skills of a data science aspirant. It covers the entire spectrum, from understanding data to deploying a model. Successful completion of this project will be a testament to the student's readiness to tackle real-world data science challenges.

**Note**: Ensure to abide by the terms and conditions of the dataset source. Always give credit to the original source when presenting or publishing your work.