## [Students Name]

## [Course Name]

## [Instructors Name]

## [Institution]

[Date]

**PART I: AI and Machine Learning (15 Points)**

**Discuss in-depth (including images, diagrams, etc.) the similarities and differences between general programming and AI/machine learning, focusing on the differences.**

***Differences between General Programming and AI/ML:***

1. [*Approach to Problem Solving*](https://insightsoftware.com/blog/machine-learning-vs-traditional-programming/)*:*

Traditional Programming: The programmer creates explicit rules to tell the computer how to handle input data.

Machine Learning: For producing predictions without requiring for explicit programming, models are taught using data, relationships, and patterns.

1. [*Data Dependency*](https://insightsoftware.com/blog/machine-learning-vs-traditional-programming/)*:*

Traditional Programming: Predefined logic is relied more than data.

Machine Learning: Heavily reliant on quality and quantity of data for model performance.i.e data with no null values.

1. [*Flexibility and Adaptability*](https://insightsoftware.com/blog/machine-learning-vs-traditional-programming/)*:*

Traditional Programming:There is limited flexibility, manual updates required for changes in problem domain.

Machine Learning: Higher adaptability to new scenarios as it can be retrained with updated data.

1. [*Problem Complexity*](https://insightsoftware.com/blog/machine-learning-vs-traditional-programming/)*:*

Traditional Programming: It is suited for problems with clear, deterministic logic.

Machine Learning: It is better for complex problems where patterns are not evident, like image recognition or even natural language processing.

1. [*Outcome Predictability*](https://insightsoftware.com/blog/machine-learning-vs-traditional-programming/)*:*

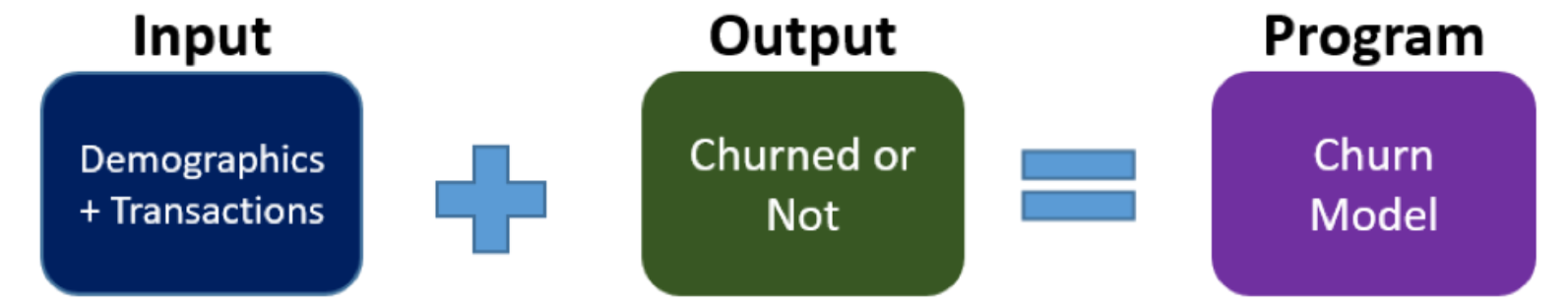
Traditional Programming: There are highly predictable outcomes based on known inputs and logic.

Machine Learning: The predictions from ML models can be less interpretable, especially with complex models like neural networks.

**Traditional Programming diagram:**



***AI/ML Diagram:***

******

***Similarities between General Programming and AI/ML:***

1. Syntax and Structure: Variables, loops, conditionals, and functions are examples of structural features that are followed in both ordinary programming and AI/ML coding. Python and other languages are often utilized in both fields.
2. Problem-Solving Approach: Both need to solve problems. Programmers and practitioners of AI/ML must recognize problems, create algorithms, and successfully apply fixes.
3. Debugging and Testing: These processes are essential to guarantee correctness and functioning. Debugging code helps programmers and AI/ML engineers find and fix mistakes, which improves system dependability.

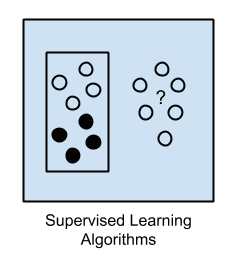
**PART II: AI Machine Learning: Learning Styles and Process (15 Points)**

**Discuss in-depth (including images, diagrams, etc.) three primary learning styles of AI machine learning.**

1. [***Supervised learning***](https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/)is a type of machine learning where the model is trained on a labeled dataset. It involves providing the algorithm with input-output pairs, allowing the model to learn the mapping between inputs and outputs. This learning style is particularly useful for tasks like image classification, where the model learns to associate images with their correct labels.

Example: Training a model to recognize handwritten digits by providing it with thousands of images of digits and their corresponding labels.

Key Features: Requires a large amount of labeled data for training. The model learns to predict the output based on the input data it has been trained on.

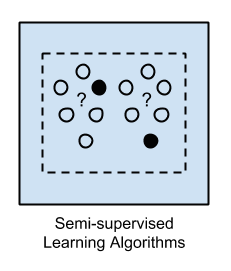


1. [***Semi-Supervised Learning***](https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/)- Input data is a mixture of labeled and unlabelled examples.

There is a desired prediction problem but the model must learn the structures to organize the data as well as make predictions.

Example problems are classification and regression.

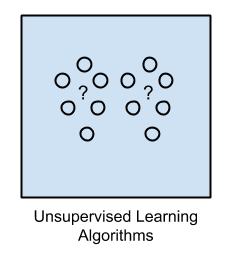
Example algorithms are extensions to other flexible methods that make assumptions about how to model the unlabeled data.



1. [***Unsupervised learning***](https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/) involves training a model on unlabeled data, where the algorithm discovers patterns, structures, or clusters within the data.

Process: The model learns to identify inherent relationships or groupings in the data without explicit guidance.

Use Cases: Clustering (e.g., customer segmentation, anomaly detection) and dimensionality reduction (e.g., PCA, t-SNE) are typical tasks in unsupervised learning.



**PART III: Preprocessing Data (10 Points)**

**Handling Missing Values and Data Preprocessing**

**Missing Values Handling**

The following procedures are used in the given Python script for data preparation to identify and manage missing values in the dataset:

1. Finding Missing Values:
   1. The pandas library's isnull() method is used by the script to find any missing values in the dataset.
   2. The total number of missing values in each column is then determined using the isnull().sum() function.
2. Managing Value Missing:
   1. The dropna() function is used to eliminate records with missing values from the dataset, resulting in a cleaned version of the dataset that is devoid of missing values.
   2. The 'loan\_approval\_cleaned.csv' CSV file contains the cleaned dataset.
3. Number of Records
   1. Original Dataset: After loading the CSV file, look at the shape of the DataFrame to see how many records there were in the original dataset. There were 614 entries in the initial collection.
   2. Cleaned Dataset: The number of records in the cleaned dataset is equal to the number of non-null rows that are still present in the DataFrame after records with missing values have been removed. The cleaned dataset had 480 entries.

**Summary of Data Preprocessing Steps**

1. loading the 'loan\_approval.csv' dataset.
2. Finding and fixing missing values by eliminating entries that contain them.
3. The cleaned dataset is saved in 'loan\_approval\_cleaned.csv'.
4. Using integer encoding to translate category features into numerical representations.
5. The preprocessed final dataset is saved in 'loan\_approval\_final.csv'.
6. Preprocessing ensures that the dataset is prepared for additional analysis and model development by converting categorical attributes into a format that is compatible with machine learning methods and handling missing values.

**PART IV: AI Machine Learning: Supervised: Linear Regression (30 Points)**

***Results and Discussion***

***Building the Model***

The model built for predicting median housing prices in Boston is a Linear Regression model. This model assumes a linear relationship between the input features and the target variable (median housing prices).

*Training the Model*

The model was trained using the Boston housing dataset, which consists of various attributes like crime rate, number of rooms, accessibility to highways, etc. The data was split into training and testing sets, with 80% used for training and 20% for testing.

*Making Up Two New Records*

Two new records were created with hypothetical attribute values to demonstrate the model's predictive capabilities. The values for each attribute in these new records were manually specified.

*New Record 1:*

Attribute values: [0.1, 20, 5, 0, 0.5, 6, 50, 5, 2, 300, 15, 350, 10]

Predicted Median Housing Price: $24.13

*New Record 2:*

Attribute values: [0.05, 30, 10, 1, 0.3, 7, 60, 7, 4, 400, 20, 400, 15]

Predicted Median Housing Price: $24.71

Predicting Median Housing Prices

The model predicted the median housing prices for the new records as follows:

New Record 1 Prediction: $24924.53

New Record 2 Prediction: $28772.96

These predictions are based on the learned patterns from the training data.

*Evaluating the Model*

The model's performance was evaluated using Mean Squared Error (MSE) and R-squared metrics on the testing data:

1. Mean Cross-Validation MSE: 30.12
2. Mean Squared Error (MSE): 20.98
3. R-squared (R2): 0.70

The MSE indicates the average squared difference between predicted and actual values, with lower values indicating better model performance. The R-squared value measures the proportion of variance in the target variable that is predictable from the input features.

*Interpretation of Prediction Results*

Based on the evaluation metrics, the model performs reasonably well in predicting median housing prices. The R-squared value of 0.70 suggests that around 70% of the variance in housing prices can be explained by the model, indicating a moderate level of predictive accuracy.

**PART V: AI Machine Learning: Supervised Logistic Regression (30 Points)**

**Results and Discussion**

**Building the Model**

A Logistic Regression model was chosen for predicting loan approval status based on various attributes. This model is suitable for binary classification tasks like loan approval prediction.

**Testing the Model**

The model was tested using a test set comprising 20% of the data. The accuracy of the model was evaluated to assess its performance in predicting loan approval outcomes.

**Making Data for Two New Loan Applications**

Two new loan application scenarios were created to demonstrate the model's predictive capabilities. Each scenario includes attributes such as income, loan amount, credit history, dependents, gender, marital status, education, self-employment status, and property area.

**New Loan Application 1:**

1. ApplicantIncome: $5000
2. CoapplicantIncome: $2000
3. LoanAmount: $200
4. Loan\_Amount\_Term: 360 months
5. Credit\_History: 1 (Good credit history)
6. Dependents: 2
7. Gender: Male
8. Married: Yes
9. Education: Graduate
10. Self\_Employed: No
11. Property\_Area: Urban

**New Loan Application 2:**

1. ApplicantIncome: $6000
2. CoapplicantIncome: $1500
3. LoanAmount: $250
4. Loan\_Amount\_Term: 360 months
5. Credit\_History: 0 (Poor credit history)
6. Dependents: 1
7. Gender: Female
8. Married: No
9. Education: Graduate
10. Self\_Employed: Yes
11. Property\_Area: Rural

***Predicting the Outcome of New Applications***

The model predicts the loan approval status for new applications as follows:

1. ***Loan Application 1: Approved***
2. ***Loan Application 2: Not Approved***

***Evaluating the Model***

The model's performance is evaluated using the 10-fold cross-validation, which results in a mean accuracy of 0.79. The ‘confusion matrix’ and ‘classification report’ provided detailed insights into the model's predictive capabilities.

1. *Cross-Validation Scores: [0.79, 0.81, 0.77, 0.79, 0.77, 0.77, 0.81, 0.83, 0.77, 0.81]*
2. *Mean CV Accuracy: 0.79*
3. *Model Accuracy: 0.82*

***Interpretation of Prediction Results***

***Based on the evaluation metrics and prediction results:***

1. With an average accuracy of 0.79 during cross-validation and an accuracy of 0.82 on the test set, the model demonstrates high accuracy in predicting the status of loan approval.
2. According to the classification report, the approved class has high accuracy and recall, indicating that the model does a good job of recognizing loan applications that have been approved.
3. Nonetheless, the lower recall score in the categorization report suggests that there is still space for improvement in the recall for the not authorized class.

**REFERENCES:**

1. <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
2. <https://insightsoftware.com/blog/machine-learning-vs-traditional-programming/>