



ITTE105b – Analytics Application

Final Term Requirement – Design, Evaluation, and Deployment of Machine Learning Models for Real-World Applications

Objectives:

The objective of this project is to develop and deploy a machine learning model that can be applied across various fields such as healthcare, education, agriculture, finance, and technology. The project involves implementing and comparing the performance of three machine learning algorithms such as Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) to identify the most effective model for the selected dataset.

Additionally, the project aims to develop, evaluate, and deploy the best-performing model through a web or mobile application interface, demonstrating how machine learning solutions can be effectively integrated into real-world applications.

I. Contents

a. Data Understanding and Preparation

- Describe the dataset, including the number of observations, features, and target variable.
- Identify data types (categorical, numerical, ordinal) and summarize key statistics for both numerical and categorical variables.
- Perform exploratory data analysis (EDA) to reveal trends, correlations, and class distributions.
- Explain all data preprocessing steps, including:
 - Handling missing values or outliers (if applicable)
 - Encoding categorical variables
 - Feature scaling or normalization
 - Train-test data splitting

b. Model Implementation

- Build three models: *Random Forest*, *KNN*, and *SVM*.
- State model parameters and tuning process (e.g., grid search, cross-validation).
- Present model training workflows, including key functions and parameters used.

c. Model Evaluation

- Evaluate model performance using metrics as *accuracy*, *precision*, *recall*, *F1-score*, and *confusion matrix*.
- Visualize results (e.g., confusion matrices, ROC curves, or performance comparison charts).
- Interpret the outcomes and discuss the models' strengths and weaknesses.

d. Comparative Analysis & Discussion

- Compare the performance of *Random Forest*, *KNN*, and *SVM* in terms of overall accuracy and performance.
- Discuss why one model might outperform others in this dataset.
- Reflect on the implications of your findings.



- Suggest potential improvements.

e. Model Deployment

- Demonstrate the deployed model showing how users can interact with the interface, submit inputs, and view the model's predictions.
- Explain the deployment process, such as:
 - How the best-performing model was selected and trained, including preprocessing, parameter tuning, and evaluation steps.
 - Saving the trained model.
 - Creating an API endpoint (e.g., using FastAPI)

Kunwari di niyo pa alam. There will be no written exam; this will serve as your final term examination. Good luck!

