

Lab07 – Lab Statement: Modeling

Course: Programming for Data Science

Objective:

This lab focuses on the **modeling phase** of the data science process, where students use prepared data to build, evaluate, and interpret predictive or descriptive models. The lab will emphasize selecting appropriate modeling techniques, implementing them programmatically, and assessing model performance in the context of business or organizational goals.

Learning Goals:

By the end of this lab, students will be able to:

1. Translate analytical goals into modeling strategies.
2. Select suitable models for classification, regression, or clustering tasks.
3. Implement models using Python and relevant libraries (e.g., scikit-learn, statsmodels).
4. Evaluate model performance using appropriate metrics and validation techniques.
5. Interpret results in a business context and discuss implications.

Lab Activities (may be highly adapted according to your previous work):

1. **Defining the Modeling Task**
 - Review the problem statement and business goals from Lab06.
 - Determine the type of modeling task (e.g., regression, classification, clustering).
 - Specify target variables, features, and expected outcomes.
2. **Feature Engineering and Data Transformation**
 - Review the cleaned and prepared dataset from Lab06.
 - Apply feature selection or dimensionality reduction techniques.
 - Create new features if necessary to improve model performance.
3. **Model Selection and Implementation**
 - Choose appropriate modeling techniques (e.g., linear regression, decision trees, k-NN, logistic regression, k-means clustering).
 - Implement models using Python, splitting the dataset into training and testing subsets.
 - Train the models and generate predictions.
4. **Model Evaluation**
 - Evaluate models using relevant metrics (e.g., accuracy, precision, recall, RMSE, silhouette score).
 - Perform cross-validation to ensure robustness.
 - Compare multiple models and justify the selection of the most effective one.
5. **Interpretation and Business Insights**
 - Analyze model outputs and feature importance.
 - Translate technical results into actionable insights for business or organizational decisions.
 - Discuss limitations and possible improvements.

Expected Output:

- **Word Document:**
 - Description of modeling task and rationale for chosen models.
 - Explanation of feature engineering and transformations.
 - Model Evaluation Results and Interpretation in a Business Context.
 - Discussion of limitations and future steps.
- **Python Code:** Implementation of models, training, evaluation, and visualization.
- **Data:** Use the cleaned dataset from Lab06 or a provided dataset for modeling.

Optional Extensions:

- Hyperparameter tuning.
- Visualization of model performance (confusion matrices, ROC curves, residual plots).
- Implementing ensemble methods (e.g., Random Forest, Gradient Boosting) for improved performance.