

Linear and Logistic Regression Assignments

Assignment 1: Basic - Predicting Session Satisfaction with Linear Regression

Objective: Build a linear regression model to predict client satisfaction scores based on session metrics for the One Hour AI Solution platform.

Task Description: As an AI engineer working with the One Hour AI Solution platform, you need to understand what factors influence client satisfaction. You'll build a linear regression model to predict satisfaction scores, helping the platform optimize the client experience.

Dataset: session_metrics.csv

Assignment Tasks:

1. Build a linear regression model to predict 'client_satisfaction_score' using at least three features
2. Split the data into training (80%) and testing (20%) sets
3. Train your model and evaluate it using R-squared and Mean Squared Error
4. Create a scatter plot showing the relationship between the most influential feature and client satisfaction
5. Create a bar chart showing the coefficients (importance) of each feature in your model

Submission Requirements:

- Python script with your code (.py file)
- Screenshots of both visualizations
- Brief explanation (2-3 sentences) of which features most strongly impact client satisfaction and why

Assignment 2: Intermediate - Problem Solvability Classification with Logistic Regression

Objective: Build a logistic regression model to predict whether an AI problem can be successfully solved in a one-hour session.

Task Description: The One Hour AI Solution platform needs to assess whether new client problems can be successfully solved within the platform's one-hour timeframe. You'll create a logistic regression model to predict success probability based on problem characteristics.

Dataset: problem_outcomes.csv

Assignment Tasks:

1. Create a 2x2 subplot dashboard with the following components:
 - Top-left: Bar chart showing the count of successful vs. unsuccessful problems
 - Top-right: Scatter plot of problem complexity vs. data size, colored by success
 - Bottom-left: Bar chart showing success rate by whether domain expertise is required
 - Bottom-right: Heatmap showing correlation between features
2. Build a logistic regression model to predict 'successfully_solved' using the available features
3. Evaluate your model using accuracy, precision, recall, and F1-score
4. Create a visualization showing feature importance in your logistic regression model

Submission Requirements:

- Python script with your code (.py file)
- Screenshots of your dashboard and feature importance visualization
- Written analysis (5-7 sentences) explaining the key factors that determine whether a problem can be solved within one hour and how this information could be used by the platform

Assignment 3: Advanced - Multi-Factor Analysis of Client Retention

Objective: Build and evaluate multiple regression models to understand what factors impact client retention probability on the One Hour AI Solution platform.

Task Description: As a data scientist for One Hour AI Solution, you need to analyze how different aspects of the client-engineer interaction affect the likelihood of client retention. You'll create sophisticated regression models to identify the most important factors.

Dataset: client_engineer_interactions.csv

Assignment Tasks:

1. Use Seaborn to create a heatmap showing the correlation between all numerical variables
2. Create a scatter plot with the following features:
 - X-axis: communication_quality
 - Y-axis: client_retention_probability
 - Point size: solution_completeness
 - Point color: follow_up_needed
 - Add a regression line to show the trend
3. Build two regression models:
 - Linear Regression model to predict client_retention_probability
 - Logistic Regression model to predict follow_up_needed (binary outcome)
4. Compare the performance of both models and identify the most influential factors in each
5. Create a custom visualization that shows how different combinations of communication_quality and solution_completeness affect retention probability

Submission Requirements:

- Python script with your code (.py file)
- Screenshots of all visualizations
- Detailed analysis (8-10 sentences) interpreting the models, identifying key patterns, and providing actionable recommendations for the One Hour AI Solution platform
- Brief explanation of which model (linear or logistic) was more effective for its respective prediction task and why