L04 Write up on Jupyter notebook

Contributors: Martin Demel, Jiri Musil, Jasmine Campos, Munachimso Benson

Department of Science, Technology, Engineering & Math, Houston Community College

ITAI - 1371 Introduction to Machine Learning

Professor Vishwa Rao

September 16th, 2025.

# Introduction

# Each team member used Jupyter Notebook to explore its features and capabilities specifically for machine learning and data science. We focused on testing code execution, data visualization, statistical analysis, and documentation functions essential for ML projects. Through hands-on testing, we discovered how Jupyter combines code, text, mathematical formulas, and interactive visualizations within a single environment. The cell-based layout, immediate feedback from running code, and rich formatting options helped us understand why Jupyter has become the preferred tool for machine learning practitioners and researchers.

# Body

We have learned that Jupyter Notebook is an interactive, web-based environment that enables us to write and run code alongside rich text documentation, visualize data instantly, and record our thought process as we analyze datasets and develop models. We verified imports for NumPy, Pandas, Matplotlib, Seaborn, and Scikit‑learn and performed basic operations.

We created a synthetic classification dataset (1000 rows, two features) and reviewed target balance and descriptive statistics. We visualized patterns with a scatter plot, histograms, and a correlation heatmap – Figure 1.

A graph of different colored bars

AI-generated content may be incorrect.

Figure 1 – Pattern visualization with scatter plot, histogram, and correlation map.

We built an ML pipeline, using a train and test split (800/200), trained Logistic Regression, and achieved 0.900 test accuracy. We also visualized the decision boundaries for both the train and test data – Figure 2.

A yellow and purple colored graph

AI-generated content may be incorrect.

Figure 2 – Training data visualization with decision boundary and predictions

We have documented ML math with LaTeX and compared algorithms, Logistic Regression (0.900), Random Forest (0.930), and SVM (0.935), identifying SVM as best – Figure 3.

A graph of different colored squares

AI-generated content may be incorrect.

Figure 3 – ML Algorithm Performance

# Conclusion

We have learned that Jupyter Notebook is more than just a coding tool. It combines code, narrative text, math, and visuals, allowing us to perform experiments, view results instantly, and understand our models step by step. We also understand that the cell-based workflow and its integration with NumPy, Pandas, Scikit‑learn, Matplotlib, and Seaborn make our work reproducible and easy to share for our Introduction to Machine Learning class.