

Name	Task	Database	Database URL
Abdul Malek Adib Mukhriz	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, ...).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Iris (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/iris">https://archive.ics.uci.edu/ml/datasets/iris</a>
Abu Lail Habib Maher	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Titanic (Kaggle)	<a href="https://www.kaggle.com/competitions/titanic">https://www.kaggle.com/competitions/titanic</a>
Adorjáni Jonathán	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or ...).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	House Prices – Ames (Kaggle)	<a href="https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data">https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data</a>
Alhijazin Katrin Sakher Ibrahim	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Wine Quality (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/wine+quality">https://archive.ics.uci.edu/ml/datasets/wine+quality</a>
Álmási Júlia	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Handle extreme class imbalance using class-weight adjustment.</li> <li>- Compare model performance before and after balancing.</li> </ul>	Credit Card Fraud (Kaggle)	<a href="https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud">https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud</a>
Amangeldin Arlan	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Adult Census Income (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/adult">https://archive.ics.uci.edu/ml/datasets/adult</a>
Amirkhan Nurgul	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Pima Indians Diabetes (Kaggle)	<a href="https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database">https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database</a>
Baikenov Zhanbolat	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use weather and temporal features.</li> <li>- Analyze seasonal patterns and evaluate models on subsets (e.g., winter vs summer).</li> </ul>	Bike Sharing (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset">https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset</a>
Bartal Dávid	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Perform data augmentation and normalization.</li> </ul>	MNIST	<a href="https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits">https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits</a>

Name	Task	Database	Database URL
Bernáth Máté Levente	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Define time-window features (e.g., hour, weekday, season).</li> <li>- Perform spatial feature engineering (pickup/dropoff clustering or ZIP-code aggregation).</li> <li>- Optionally model separate subsets (weekdays vs weekends).</li> </ul>	NYC Taxi (TLC)	<a href="https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page">https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page</a>
Blilkevicius Motiejus	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use sliding windows and create multi-step forecasting targets.</li> <li>- Experiment with resampling (hourly/daily averages).</li> </ul>	ElectricityLoadDiagrams20112014	<a href="https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014">https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014</a>
Bobkov Aleksei	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Iris (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/iris">https://archive.ics.uci.edu/ml/datasets/iris</a>
Chang Jia Qian	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Titanic (Kaggle)	<a href="https://www.kaggle.com/competitions/titanic">https://www.kaggle.com/competitions/titanic</a>
Cory Jonathan David	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	House Prices – Ames (Kaggle)	<a href="https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data">https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data</a>
Egorov Andrey	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Wine Quality (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/wine+quality">https://archive.ics.uci.edu/ml/datasets/wine+quality</a>
Fodor Kornél Máté	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Handle extreme class imbalance using SMOTE, undersampling, or class-weight adjustment.</li> <li>- Compare model performance before and after balancing.</li> </ul>	Credit Card Fraud (Kaggle)	<a href="https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud">https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud</a>
Gong Jiayu	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Adult Census Income (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/adult">https://archive.ics.uci.edu/ml/datasets/adult</a>
Hennig Luca Silvio	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Pima Indians Diabetes (Kaggle)	<a href="https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database">https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database</a>

Name	Task	Database	Database URL
Khudiyeva Aydan	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use weather and temporal features.</li> <li>- Analyze seasonal patterns and evaluate models on subsets (e.g., winter vs summer).</li> </ul>	Bike Sharing (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset">https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset</a>
Kiss Boróka	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Perform data augmentation and normalization.</li> </ul>	MNIST	<a href="https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits">https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits</a>
Kossuth Hugó Ádám	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Define time-window features (e.g., hour, weekday, season).</li> <li>- Perform spatial feature engineering (pickup/dropoff clustering or ZIP-code aggregation).</li> <li>- Optionally model separate subsets (weekdays vs weekends).</li> </ul>	NYC Taxi (TLC)	<a href="https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page">https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page</a>
Margitics László	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use sliding windows and create multi-step forecasting targets.</li> <li>- Experiment with resampling (hourly/daily averages).</li> </ul>	ElectricityLoadDiagrams20112014	<a href="https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014">https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014</a>
Mátay-Csepeli Gábor András	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Iris (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/iris">https://archive.ics.uci.edu/ml/datasets/iris</a>
Matesz Réka	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Titanic (Kaggle)	<a href="https://www.kaggle.com/competitions/titanic">https://www.kaggle.com/competitions/titanic</a>
Medvedev Matvei	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	House Prices – Ames (Kaggle)	<a href="https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data">https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data</a>
Medvegy Gábor	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Wine Quality (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/wine+quality">https://archive.ics.uci.edu/ml/datasets/wine+quality</a>
Miha Anna	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Handle extreme class imbalance using SMOTE, undersampling, or class-weight adjustment.</li> <li>- Compare model performance before and after balancing.</li> </ul>	Credit Card Fraud (Kaggle)	<a href="https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud">https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud</a>

Name	Task	Database	Database URL
Nagyhegyesi Luca	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Adult Census Income (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/adult">https://archive.ics.uci.edu/ml/datasets/adult</a>
Nagy Mihály Gyula	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Pima Indians Diabetes (Kaggle)	<a href="https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database">https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database</a>
Nagy Tamás	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use weather and temporal features.</li> <li>- Analyze seasonal patterns and evaluate models on subsets (e.g., winter vs summer).</li> </ul>	Bike Sharing (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset">https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset</a>
Nguyen Hong Gia Bao	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Implement a Convolutional Neural Network (CNN) as the main model.</li> <li>- Compare CNN performance with a classical ML model such as SVM.</li> <li>- Perform data augmentation and normalization.</li> </ul>	MNIST	<a href="https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits">https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits</a>
Nguyen Thanh Trung	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Define time-window features (e.g., hour, weekday, season).</li> <li>- Perform spatial feature engineering (pickup/dropoff clustering or ZIP-code aggregation).</li> <li>- Optionally model separate subsets (weekdays vs weekends).</li> </ul>	NYC Taxi (TLC)	<a href="https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page">https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page</a>
Onder Sára	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use sliding windows and create multi-step forecasting targets.</li> <li>- Experiment with resampling (hourly/daily averages).</li> </ul>	ElectricityLoadDiagrams20112014	<a href="https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014">https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014</a>
Péché Kornél	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Iris (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/iris">https://archive.ics.uci.edu/ml/datasets/iris</a>
Schmiedt Lilla	<p>Project Type: Classification.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Titanic (Kaggle)	<a href="https://www.kaggle.com/competitions/titanic">https://www.kaggle.com/competitions/titanic</a>
Sólyom Livia	<p>Project Type: Regression.</p> <p>Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	House Prices – Ames (Kaggle)	<a href="https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data">https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data</a>

Name	Task	Database	Database URL
Somlai Dóra	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Wine Quality (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/wine+quality">https://archive.ics.uci.edu/ml/datasets/wine+quality</a>
Szabó Bence	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Handle extreme class imbalance using SMOTE, undersampling, or class-weight adjustment.</li> <li>- Compare model performance before and after balancing.</li> </ul>	Credit Card Fraud (Kaggle)	<a href="https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud">https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud</a>
Szugiczki Flóra	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Adult Census Income (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/adult">https://archive.ics.uci.edu/ml/datasets/adult</a>
Tóth Borbála Vivien	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul>	Pima Indians Diabetes (Kaggle)	<a href="https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database">https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database</a>
Török Zsófia	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use weather and temporal features.</li> <li>- Analyze seasonal patterns and evaluate models on subsets (e.g., winter vs summer).</li> </ul>	Bike Sharing (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset">https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset</a>
Tulit Zolt	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Perform data augmentation and normalization.</li> </ul>	MNIST	<a href="https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits">https://archive.ics.uci.edu/dataset/683/mnist+database+of+handwritten+digits</a>
Tursynbay Zhaniya	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Define time-window features (e.g., hour, weekday, season).</li> <li>- Perform spatial feature engineering (pickup/dropoff clustering or ZIP-code aggregation).</li> <li>- Optionally model separate subsets (weekdays vs weekends).</li> </ul>	NYC Taxi (TLC)	<a href="https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page">https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page</a>
Tussipbek Madi	<p>Project Type: Regression. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Linear Regression and a Decision Tree Regressor.</li> <li>- Train at least one advanced model (e.g., Boosting Regressor, Random Forest Regressor).</li> <li>- Provide evaluation using: MAE, RMSE, <math>R^2</math> score, and residual-analysis plots.</li> </ul> <p>Additional Requirements:</p> <ul style="list-style-type: none"> <li>- Use sliding windows and create multi-step forecasting targets.</li> <li>- Experiment with resampling (hourly/daily averages).</li> </ul>	ElectricityLoadDiagrams20112014	<a href="https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014">https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014</a>
Zádor Levente	<p>Project Type: Classification. Your task is to build a complete machine-learning pipeline using the selected dataset. This should include exploratory data analysis, data cleaning, feature engineering, model training, hyperparameter tuning, and evaluation. Provide clear visualizations that support your decisions.</p> <ul style="list-style-type: none"> <li>- Train at least two baseline models, including Logistic Regression and a Decision Tree.</li> <li>- Train at least one advanced model (e.g., Random Forest, Boosting, or SVM).</li> <li>- Provide evaluation using: confusion matrix, classification report, ROC curve.</li> </ul>	Iris (UCI)	<a href="https://archive.ics.uci.edu/ml/datasets/iris">https://archive.ics.uci.edu/ml/datasets/iris</a>