Final Project

Machine Learning 2

Author: Lina Faik

Date: May 2025

Table of contents

01 Objectives

02 Project Ideas

03 Guidelines

04 Evaluation Criteria

05 Q&A

01

Objectives

01. Objectives

✓ **Apply Machine Learning Techniques:**

Utilize advanced machine learning algorithms learned in the course to solve real-world problems.

✓ **End-to-End Project Execution:**

Experience the entire machine learning workflow from data collection, preprocessing, and model training to evaluation and deployment.

✓ **Critical Thinking and Problem Solving:**

Develop the ability to identify, analyze, and solve complex problems using data-driven approaches.

✓ **Collaborative Work:**

Enhance teamwork and project management skills by working effectively in groups.

✓ **Communication Skill:**

Improve the ability to communicate technical results and insights through reports and presentations.

✓ **Practical Implementation:**

Gain hands-on experience with tools and libraries used in the industry, such as Python, scikit-learn, TensorFlow, or similar frameworks.

02

Project Ideas

Overview

Team Formation:

- Form teams of 3-4 members.
- Individual projects are not allowed.

Project Selection:

- Choose a project from the provided list.
- You are also welcome to propose a different project topic if it aligns more with your interests.

Objectives and Data Sources:

- The listed objectives and data sources are suggestions.
- You are responsible for defining your project's specific problem statement and validating the relevance of your data sources.

Overview

Summary of the projects.

Provided as a guideline. Specifics will depend on your chosen approach and tools.

	Difficulty (1-5)	Topics
Project 1: Sentiment Analysis on Social Media	2/5	NLP
Project 2: Customer Segmentation	3/5	Clustering
Project 3: Recommendation System for E-commerce	4/5	NLP
Project 4: Stock Price Prediction	3/5	Time Series
Project 5: Predicting Customer Churn	2/5	Clustering, ML
Project 6: Predictive Maintenance	3/5	ML, Time Series
Project 7: Loan Default Prediction	2/5	ML
Project 8: Music Genre Classification	4/5	ML / DL
Project 9: Predicting Employee Attrition	2/5	ML
Project 10: Energy Consumption Forecasting	3/5	Time Series

Project 1: Sentiment Analysis on Social Media

Context:

- Social media platforms are a rich source of public opinion.
- Understanding sentiment can help brands and organizations gauge public perception.
- Real-time sentiment analysis can inform marketing and communication strategies.

Goal:

- Develop a model to classify the sentiment of social media posts (positive, negative, neutral).
- Analyze trends in sentiment over time or across different topics.
- Visualize the results to provide actionable insights.

Data:

- Kaggle Twitter US Airline Sentiment
- Sentiment140
- Social Media Sentiment Data

Challenges:

- Preprocessing text data (removing noise, handling slang, emojis).
- Balancing the dataset if there is an uneven distribution of sentiment classes.
- Managing the large volume of data and ensuring the model scales effectively.

Project 2: Customer Segmentation

Context:

- Businesses need to understand their customer base to tailor marketing strategies.
- Segmentation helps in identifying groups with similar behaviors and preferences.
- Data-driven segmentation can improve customer satisfaction and retention.

Goal:

- Use clustering techniques to segment customers based on purchasing behavior.
- Analyze the characteristics of each segment.
- Provide recommendations for targeted marketing campaigns.

Data:

- Kaggle Customer Segmentation
- Mall Customer Segmentation Data
- Online Retail Dataset

Challenges:

- Determining the optimal number of clusters.
- Interpreting the clusters in a business context.
- Ensuring the model scales with large datasets.

Project 3: Recommendation System for E-commerce

Context:

- E-commerce platforms benefit from personalized recommendations to enhance user experience.
- Recommendations can increase sales and user engagement.
- Machine learning can automate and improve the recommendation process.

Goal:

- Build a recommendation system to suggest products to users based on their behavior.
- Compare collaborative filtering and content-based approaches.
- Evaluate the effectiveness of the recos using metrics like precision and recall.

Data:

- Kaggle Amazon Product Data
- MovieLens Dataset
- Goodreads Books

Challenges:

- Handling sparse data in user-item interaction matrices.
- Scaling the recommendation system for large datasets.
- Balancing diversity and relevance in recommendations.

Project 4: Stock Price Prediction

Context:

- Financial markets are influenced by numerous factors, making prediction challenging.
- Accurate predictions can guide investment strategies.
- Historical data can be analyzed to identify patterns and trends.

Goal:

- Develop a model to predict future stock prices based on historical data.
- Explore different time series forecasting methods.
- Evaluate the model's performance and identify potential risks.

Data:

- Yahoo Finance Historical Data

Challenges:

- Handling noisy and non-stationary data.
- Incorporating external factors like news and economic indicators.
- Preventing overfitting to historical trends.

Project 5: Predicting Customer Churn

Context:

- Customer retention is crucial for business profitability.
- Identifying factors leading to churn can help in devising retention strategies.
- Machine learning can automate the identification of at-risk customers.

Goal:

- Build a model to predict whether a customer will churn based on their behavior.
- Analyze the features that contribute most to churn.
- Provide recommendations for reducing churn.

Data:

- Kaggle Telco Customer Churn
- Customer Churn Data

Challenges:

- Balancing the dataset if there is an uneven distribution of churned vs. retained customers.
- Feature engineering to capture relevant customer behaviors.
- Interpreting the model to provide actionable business insights.

Project 6: Predictive Maintenance

Context:

- Equipment failures can lead to costly downtime and repairs.
- Predictive maintenance aims to predict failures before they occur.
- Machine learning can analyze sensor data to identify early signs of failure.

Goal:

- Develop a model to predict equipment failures based on sensor data.
- Identify the most critical factors contributing to failures.
- Provide recommendations for maintenance schedules.

Data:

- NASA Turbofan Engine Degradation Simulation Data Set
- Predictive Maintenance Dataset
- Predictive Maintenance Dataset 2

Challenges:

- Handling large volumes of high-dimensional sensor data.
- Balancing the dataset if failures are rare.
- Interpreting the model to provide actionable maintenance insights.

Project 7: Loan Default Prediction

Context:

- Financial institutions need to assess the risk of loan defaults.
- Accurate predictions can help in making informed lending decisions.
- Machine learning can analyze historical data to predict defaults.

Goal:

- Build a model to predict whether a loan will default based on applicant data.
- Identify the most significant factors contributing to defaults.
- Provide recommendations for risk mitigation.

Data:

- Kaggle Lending Club Loan Data
- Loan Default Data
- Home Credit Default Risk

Challenges:

- Handling imbalanced datasets where defaults are less frequent.
- Feature engineering to capture relevant applicant characteristics.
- Interpreting the model to provide actionable insights.

Project 8: Music Genre Classification

Context:

- Music streaming services can benefit from automatic genre classification.
- It enhances user experience by enabling better music recommendations.
- Audio features can be analyzed to classify genres.

Goal:

- Develop a model to classify music tracks into genres based on audio features.
- Explore different feature extraction techniques.
- Evaluate the model's performance and interpret the results.

Data:

- GTZAN Music Genre Dataset
- FMA: A Dataset For Music Analysis
- Million Song Dataset

Challenges:

- Extracting relevant features from audio data.
- Handling class imbalances if certain genres are less represented.
- Ensuring the model generalizes well to new and diverse music tracks.

Project 9: Predicting Employee Attrition

Context:

- Employee attrition can impact organizational productivity and morale.
- Predicting attrition can help in retaining valuable employees.
- Machine learning can analyze employee data to identify attrition risks.

Goal:

- Build a model to predict whether an employee will leave the company.
- Analyze the factors contributing most to attrition.
- Provide recommendations for improving employee retention.

Data:

- IBM HR Analytics Employee Attrition & Performance
- Employee Attrition Data
- HR Analytics Dataset

Challenges:

- Handling imbalanced datasets if attrition is less frequent.
- Feature engineering to capture relevant employee characteristics.
- Interpreting the model to provide actionable insights for HR practices.

Project 10: Energy Consumption Forecasting

Context:

- Accurate energy consumption forecasting can help in managing resources.
- It is crucial for energy providers to balance supply and demand.
- Machine learning can analyze historical data for accurate forecasting.

Goal:

- Develop a model to forecast future energy consumption based on historical data.
- Identify patterns and trends in energy usage.
- Evaluate the model's performance and provide recommendations.

Data:

- UCI Energy Consumption Dataset
- UK National Grid Data
- Household Electricity Consumption

Challenges:

- Handling Seasonality and Trends: Capturing daily and seasonal consumption patterns.
- Incorporating External Factors: Integrating weather, holidays, and economic activity data.
- Dealing with Data Quality: Managing missing values and anomalies in historical data.

03

Guidelines

03. Guidelines

Project Planning:

- Define clear objectives and deliverables.
- Create a project timeline with milestones.

Data Handling:

- Identify and validate data sources.
- Clean and preprocess data, handling missing values.

Model Development:

- Choose suitable machine learning algorithms.
- Perform feature engineering and model training.

Evaluation:

- Use appropriate evaluation metrics.
- Implement cross-validation to ensure robustness.

Insights:

- Understand and explain model predictions.
- Derive practical insights and recommendations.

Collaboration:

- Assign team roles and responsibilities.
- Hold regular team meetings and maintain documentation.

03. Guidelines

Presentation:

- Create clear visualizations.
- Prepare a final report and presentation.

Resources:

- Use tools like Python, scikit-learn, TensorFlow.
- Refer to course materials and online resources.

Feedback:

- Seek peer and instructor feedback.
- Iterate on your model and approach as needed.

04

Evaluation Criteria

04. Evaluation Criteria

1. Project Planning and Execution (10%):

- Clarity and feasibility of project objectives.
- Adherence to the project timeline and milestones.
- Effective project management and teamwork.

2. Data Handling (20%):

- Quality and relevance of the data sources used.
- Effectiveness of data preprocessing and cleaning.
- Appropriateness of feature engineering.

3. Model Development (20%):

- Selection and justification of machine learning algorithms.
- Model training and tuning processes.
- Comparison and performance of different models.

4. Evaluation and Validation (20%):

- Use of appropriate evaluation metrics.
- Implementation of cross-validation techniques.
- Robustness and generalization of the model.

04. Evaluation Criteria

5. Insights and Interpretation (10%):

- Clarity in explaining model predictions.
- Derivation of actionable insights and recommendations.
- Understanding of feature importance and model behavior.

6. Presentation and Reporting (10%):

- Quality and clarity of visualizations.
- Comprehensive and well-structured final report.
- Effectiveness of the final presentation.

7. Ethical Considerations (5%):

- Addressing potential biases in data and models.

8. Innovation and Creativity (5%):

- Originality and creativity in problem-solving.
- Implementation of innovative techniques or approaches.

Q&A

Thank you!