Final Project – Machine Learning Course
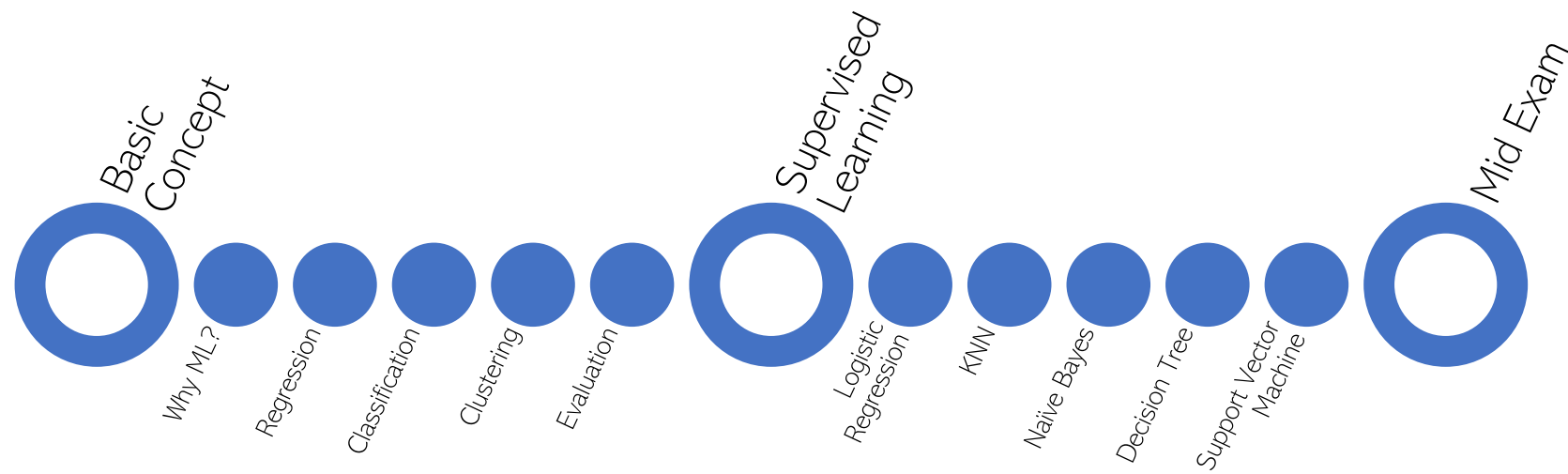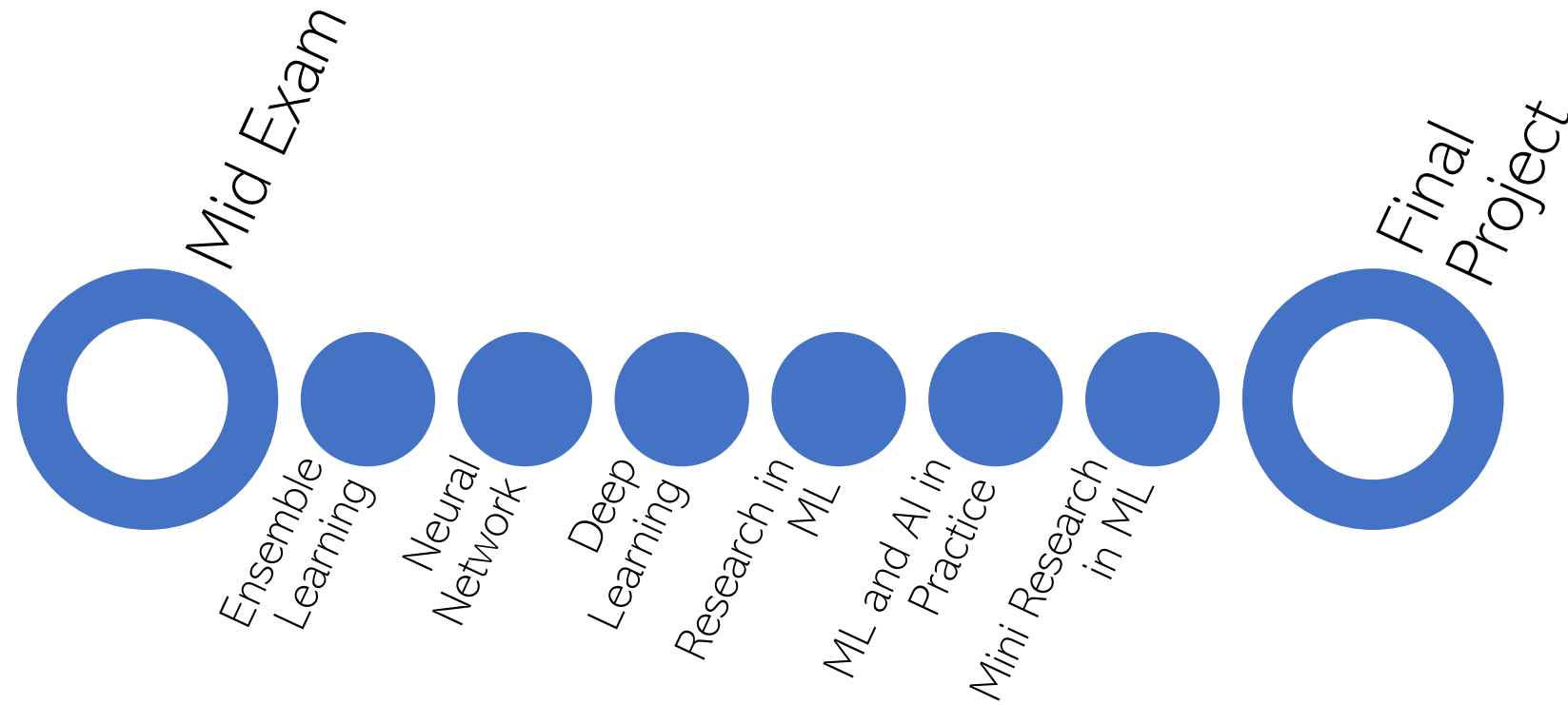
Dr. Ariana Yunita

Program Studi Ilmu Komputer
Universitas Pertamina

Semester Genap 2024 2025

Universitas
Pertamina

Mid Exam

Final Project

Ensemble Learning

Neural Network

Deep Learning

Research in ML

ML and AI in Practice

Mini Research in ML

AGENDA

01
Goals and Description

02
Problem Statement

Goals
- ✓ To allow students to apply the concepts, algorithms, and techniques learned throughout the course (e.g., supervised/unsupervised learning, model evaluation).
- ✓ To challenge students to define or/and solve real-world problems using machine learning methods.
- ✓ To expose students to the full ML workflow: Data collection & cleaning, Feature engineering, Model selection and training, Hyperparameter tuning, Model evaluation, Interpretation and reporting
- ✓ To motivate students to explore advanced topics or novel applications.
- ✓ To enhance teamwork (in group projects) and the ability to communicate technical work clearly via reports and presentations.
- ✓ To help students create a tangible project that can be showcased in their portfolio or resume.

Description

- ✓ We recommend 4-5 persons in a group
- ✓ Output from this Final Project : source code, manual book, preprocessed dataset, paper
- ✓ Presentation scheduled on Tuesday, 17 June 2025 at 10.00- 14.00 and Wednesday 18 June 2025 at 10.00-12.00
- ✓ All the documents should be submitted on 1 July 2025 via E-Learning

- ✓ Grading Indicator
- ✓ - Each group should implement a proposed model and compare with other baseline models

01 Goals and Description

02 **Problem Statement**

Case 1 : Employing Machine Learning for the Detection of Indonesian Regional Accents

Background

Indonesia is characterized by rich ethnic diversity, with each group possessing distinct cultural identities—one of which is reflected in their spoken accents. An individual's ethnic background can often be inferred from their manner of speaking, as these accents, though varying across ethnic groups, exhibit consistent and identifiable patterns. These speech patterns form a crucial part of linguistic identity and hold potential for various applications such as speaker recognition, dialect analysis, and the personalization of text-to-speech systems. Despite this potential, research on the identification and classification of Indonesian accents remains limited and underexplored.

Case 2 : Deep Learning Model for Supporting Green Economy in Indonesia

Background

Indonesia is moving forward with an energy transition focused on sustainability and environmental responsibility. The government has outlined ambitious goals to raise the proportion of New and Renewable Energy (NRE) in the national energy mix to 23% by 2025 and 31% by 2050. Reaching these milestones will require substantial investment in green energy infrastructure, such as solar, wind, hydroelectric, and geothermal power.

Despite the encouraging growth trends in the green energy sector, volatility in the stock prices of companies operating in this domain, combined with market instability and various external factors, continues to pose significant challenges to investor engagement. As a result, the inflow of capital into sustainable energy initiatives remains limited, potentially delaying the realization of the national renewable energy targets.

Case 3 : Sentiment Analysis using Machine Learning Approach for Supporting Green Economy in Indonesia

Background

Indonesia is striving to enhance the share of renewable energy within its national energy mix as part of its commitment to sustainable development and climate goals. While government policies and investment strategies play a central role in achieving this target, public perception also holds significant influence in shaping the direction and pace of renewable energy adoption. In the digital era, social media has become a dynamic platform where individuals express their opinions, concerns, and level of awareness regarding various issues, including environmental and energy-related topics.

Analyzing public sentiment toward renewable energy through social media platforms—such as Twitter, Facebook, or Instagram—offers a rich and real-time source of data that can reflect societal attitudes and potential acceptance or resistance to energy transitions. These user-generated insights provide valuable feedback for stakeholders, policymakers, and energy companies, helping them to design more responsive communication strategies, identify public misconceptions, and enhance community engagement. As such, leveraging social media data through text mining and sentiment analysis can support evidence-based decision-making and strengthen the implementation of renewable energy initiatives in Indonesia.

Case 4 : Analyzing Sentiments Toward Herbal Treatments for Diabetes: A Machine Learning Approach

Background
Diabetes mellitus is one of the most prevalent chronic diseases worldwide, with rising incidence rates particularly in low- and middle-income countries. The long-term complications associated with diabetes have prompted patients to seek alternative and complementary therapies alongside conventional medical treatments. Among these alternatives, herbal medicine has garnered significant attention due to its perceived natural origin, affordability, and accessibility.

In Indonesia and many other countries with rich traditions in herbal remedies, public discourse around the use of herbal medicine for managing diabetes has been growing, especially on digital platforms such as social media, forums, and review websites. These user-generated contents provide a wealth of information that reflects public perceptions, experiences, and levels of trust toward herbal treatments. However, while clinical studies on the efficacy of certain herbal remedies exist, there is limited research on how the public perceives these treatments, particularly in terms of sentiment.

Sentiment analysis, a subfield of natural language processing (NLP), enables the systematic extraction of opinions, attitudes, and emotions from textual data. By applying machine learning techniques to analyze sentiment in online discussions about diabetes and herbal medicines, researchers can uncover insights into public attitudes, concerns, and preferences. These insights are valuable for healthcare practitioners, policymakers, and pharmaceutical companies in understanding public behavior and designing more targeted educational or intervention programs.

Case 5 : Sentiment Analysis on Universities under APERTI BUMN Alliance through Google Review

Background:

In the era of digital transformation, online platforms have become a significant medium through which the public expresses opinions and evaluations of various institutions, including higher education. Among these platforms, Google Review stands out as a prominent source of user-generated content, offering valuable perspectives on institutional quality, service delivery, academic environment, and overall student experience.

As members of the APERTI BUMN consortium—comprising universities affiliated with Indonesian state-owned enterprises—these institutions occupy a strategic position in fostering high-quality human capital to meet national development goals. Understanding public sentiment toward these universities is therefore essential for continuous institutional improvement and informed decision-making.

Sentiment analysis, particularly through machine learning approaches, presents a robust method for systematically extracting and categorizing subjective information from large volumes of textual data. By leveraging such methods, it becomes possible to identify prevalent patterns of opinion, uncover areas of concern, and support data-driven policy and service enhancements. Despite the growing importance of sentiment analysis in the higher education sector, studies focusing on Indonesian university reviews, especially those concerning APERTI BUMN members, remain scarce and underexplored.

Case 6 : Image Waste Classification on Trashnet Dataset

Background:

Rapid urbanization, population growth, and industrialization have become major contributors to environmental pollution across the globe. The increase in production and marketing activities resulting from these processes has led to the intensive exploitation of natural resources. At the same time, rising consumption trends have generated large quantities of waste with hazardous contents, posing significant threats to both human health and the environment. Waste can be categorized based on several factors such as consumption sources, production processes, and chemical and physical characteristics. Therefore, systematic waste management is essential for safeguarding public health and environmental sustainability. Failure to prioritize recycling efforts not only leads to economic losses but also accelerates the depletion of finite natural resources.

To mitigate the negative impacts of waste, recycling is considered an effective solution. However, the success of recycling initiatives depends heavily on the ability to accurately separate waste based on its type. Traditional manual sorting methods have several limitations, including inconsistency, high labor requirements, and a high potential for human error.

With advances in artificial intelligence (AI) and digital image processing, automated approaches to waste classification have become increasingly relevant. By leveraging machine learning and deep learning techniques, image-based classification systems can efficiently and accurately recognize and categorize different types of waste, such as organic, inorganic, plastic, metal, and glass. This approach not only enhances the effectiveness of waste management systems but also supports sustainable environmental practices.
This study aims to develop an image classification model for waste that can serve as an intelligent solution for modern waste management systems, contributing to more efficient recycling practices and environmental preservation.

Table 1. Mini Research Topic, Dataset, Min-Max Number of Groups

No	Topic	Dataset	Min-Max No of Groups	Details
1	Employing Machine Learning for the Detection of Indonesian Regional Accents	V	1	
2	Deep Learning Model for Supporting Green Economy in Indonesia	-	2-3	
3	Sentiment Analysis using Machine Learning Approach for Supporting Green Economy in Indonesia	V	1-2	
4	Analyzing Sentiments Toward Herbal Treatments for Diabetes: A Machine Learning Approach	V	1-2	
5	Sentiment Analysis on Universities under APERTI BUMN Alliance through Google Review	V	1-2	
6	Image Waste Classification on Trashnet Dataset	V	1-2	

Terima Kasih