Machine learning for Data Engineers

Session 5: Project Topics

Stephany Rajeh, Ph.D.

Ecology, Climate Change, Animal Welfare, and Circular Economy

→ Each group can choose one or more topics from the following : Ecology, climate change, animal welfare, &/or circular economy.
→ Each group can also choose one of the first three (ecology, climate change, & animal welfare) & relate it to circular economy

Ecology: Predicting Deforestation or Forest Fire Risks

A) Possible project ideas:

- **Predict Deforestation Rates:** Train a model to predict future deforestation areas based on features such as satellite imagery, land use, proximity to roads, and economic activities.
- **Forest Fire Risk Prediction:** Use historical weather data and vegetation data to predict the likelihood of forest fires in specific regions.
- **Species Habitat Prediction:** Predict areas where endangered species are most likely to thrive based on environmental features.
- Etc…

B) Datasets:

- Global Forest Watch - Deforestation and forest cover datasets.
- MODIS Fire and Thermal Anomalies Data - Satellite fire data.
- WorldClim - Global climate data for species modeling.
- Copernicus Land Monitoring Service - Provides land cover data including forests, urban areas, and agricultural land.
- Inventaire Forestier National – IGN - Data on forest types, density, and health across France.
- Prométhée Database - Contains forest fire records in France, including burned areas, causes, and dates.

Climate Change: Carbon Emissions Analysis

A) Possible project ideas:

- **Carbon Footprint Estimation:** Predict the carbon emissions of industries or countries based on their economic and energy consumption patterns.
- **Climate Change Trends Analysis:** Use time-series analysis on temperature, CO2 concentration, and sea-level datasets to predict future climate change trends.
- **Renewable Energy Potential Prediction:** Build models to assess the potential of solar, wind, or hydroelectric power generation for specific regions.
- Etc…

B) Datasets:

- Climate Data Store - European climate data.
- Carbon Dioxide Information Analysis Center - CO2 emissions datasets.
- Open Power System Data - Renewable energy datasets.
- RTE Open Data - Data from France's electricity grid, including renewable energy generation (solar, wind, hydro) by region.
- Météo France - Historical and real-time weather data, including temperature, precipitation, and wind patterns.

Animal Welfare: Wildlife Monitoring and Behavior Prediction

A) Possible project ideas:

- **Animal Population Estimation:** Train a model to estimate populations of specific animals using drone imagery, thermal images, or audio recordings.
- **Animal Activity Recognition:** Classify animal behaviors (e.g., foraging, resting, moving) using sensor or video data.
- **Illegal Poaching Detection:** Detect poaching activities using spatial data and patterns of reported poaching incidents.
- Etc …

B) Datasets:

- iNaturalist - Biodiversity observations.
- Global Biodiversity Information Facility (GBIF) - Open biodiversity data.
- Wildlife Insights - Camera trap images for wildlife.
- Vigie Nature - Citizen science data on bird populations across France.

Circular Economy: Waste Management Optimization

A) Possible project ideas:

- **Recycling Rate Prediction:** Predict recycling rates of municipalities based on socio-economic and policy data.
- **Trash Sorting with Computer Vision:** Train a convolutional neural network (CNN) to classify types of trash (e.g., plastic, paper, metal) for automated waste management systems.
- **Optimal Circular Economy Strategies:** Use optimization techniques to propose circular economy strategies (e.g., material reuse, repair, recycling) for industries.
- Etc …

B) Datasets:

- Kaggle Waste Classification Dataset - Trash classification images.
- OECD Environmental Data - Waste and recycling statistics.
- Waste Management Dataset - Open datasets on waste management processes.
- EU Waste Dataset - Data on recycling rates, waste generation, and management strategies.
- ADEME (Agence de la transition écologique) – French reports and datasets on waste streams, including packaging, food waste, and e-waste.
- Data Sud - Data on waste management and recycling activities in the southern regions of France.

Multidisciplinary: Predicting Species Migration due to Climate Change

A) Possible project ideas:

- **Species Migration Patterns:** Predict how species' habitats will shift due to changes in temperature and precipitation using climate data and biodiversity records.
- **Impact of Urbanization on Biodiversity:** Analyze how urban expansion affects species diversity and suggest mitigation strategies.
- Etc …

B) Datasets:

- CHELSA Climate Data - High-resolution climate data for ecological modeling.
- NASA Earth Data - Climate and remote sensing data.
- French Land Use Data (BD TOPO) - Detailed maps of urban development across France.

Advanced Challenge: Predicting Renewable Energy Storage Needs

A) Possible project ideas:

- **Energy Demand and Storage Forecasting:** Predict renewable energy storage needs for a region based on energy generation patterns, weather forecasts, and consumption trends.
- **Battery Health Prediction:** Build a time-series model to predict the lifespan and efficiency of batteries used in renewable energy systems.
- Etc …

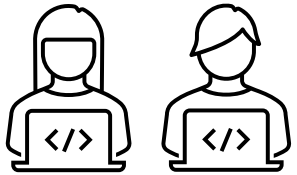
B) Datasets:

- National Renewable Energy Laboratory (NREL) - Renewable energy datasets.
- Open Energy Platform - Datasets on energy usage and generation.
- Energy Information Administration (EIA) - Global energy data.
- Photovoltaic Geographical Information System (PVGIS) - Solar and wind energy potential datasets.
- IRENA (International Renewable Energy Agency) - Reports and datasets on energy storage technologies in France and Europe.

Deliverables

Each group has 15 minutes maximum, to present the research presentation based on the outline shared “ProjectOutline.pdf”:

1. **PPT to be uploaded in a shared folder on Teams**
2. **Code to be made public on GitHub**



Tip: Start with simpler models (linear regression, random forests) and then build up to deep learning techniques