Link to Github repository for ML projects

Contract with the edition society (in English) | Amazon link

Book: DUCRAY, Joseph, 140 ans plus tard, Paris, BoD Editions

Context

- As an avid reader, I have immersed myself in nearly one book a week since 2020. This sparked a profound passion for writing. 140 Years Later was written during my first three years at university, developed in my spare time over the course of two years.
- **Key motivators**: my heritage and my determination to explore what is often regarded as 'unsolvable'
- The book is structured around **100 chronologically ordered questions**, designed to provide factual insights, with a chronological timeframe for each chapter.
- This format allows readers to consult specific topics with ease. While the book is predominantly factual, I address the my opinion on bias in the introduction.
- It is **not a solution,** nor an essay, but an **invitation** to **reflect deeply** of often **oversimplified subjects**.

Key Learnings:

1. **The Limitations of Artificial Intelligence**
- Platforms like ChatGPT can provide helpful summaries but cannot replace the depth and nuance offered by books. Each question in my book could form the basis of an entire volume—highlighting the inadequacy of quick answers for such multifaceted issues.
2. **Difference between listening and hearing.**
- Hearing is passive, listening is active. Before expressing a firm opinion, I learned to imagine presenting it to someone who has endured personal tragedy. This mental exercise fosters sensitivity and reflexion in any discussions, for any context.
  What if tomorrow we confront of the leading member from our opposite party?
3. **Understanding Multi-layered Problems**
- The conflicts in the Middle East are not monolithic; an economical rivalry may be caused by a geopolitical rivalry which may be caused by a religious one. This layered complexity is analogous to statistical covariates, where seemingly independent variables are deeply interlinked, and one can be the cause for the other.

Beginning of the introduction:

"The question is asked: who can convince a mother who lost a child?
'My wildest dream would be to restore meaning to the word,' said the other. This phrase, beyond its beauty, seeks to convey that in a time where reflection has been replaced by impulse, the impact of the word, due to the erosion of its meaning, loses its value. For the subject we are about to address, however, it is essential to recognize that words carry weight. Two terms, which reflect the same tangible realities, can represent vastly different philosophical, political, or ideological opinions.

Flood Risk Prediction and Visualization Tool

Context:

Flooding poses significant risks to the UK, exacerbated by climate change. This project aims to develop a Python-based tool for classifying flood risk, estimating house prices, and visualizing rainfall and river data. The tool provides rapid, data-driven assessments of flood risk using machine learning and integrates historical, geographic, and real-time environmental data. The project was completed in a demanding five-day period.

Key Objectives

1. **Flood Risk Prediction:** Classifying flood-prone areas to support mitigation strategies.
2. **House Price Regression:** Estimating property values in the context of environmental risk.
3. **Historic Flood Classification:** Identifying and analysing past flood events.
4. **Risk Estimation:** Delivering robust predictions for areas without labeled data.
5. **Data Visualization:** Enhancing understanding through interactive visual tools.

Key learning

- Managing multiple files, functions, and classes presented **significant debugging challenges**. Repeated testing and troubleshooting were required to ensure seamless integration of models and visualization components.
- Labeled data with precise UK postcodes required robust predictions for locations without labels. To avoid overfitting and ensure model generalizability, **geographic and temporal data splitting techniques were employed.**
- Used **XGBoost** and **Random Forest,** both regressor and classifier. Tuned the models, investigating various sklearn function: Grid Search, Cross validation, SMOTE for handling unbalanced data.
- A **thorough understanding** of the two **machine learning algorithms** and their **parameters** was critical to justify their use in the final solution
- Limitations of Machine Learning Models: Recognized that even **well-tuned** models may **perform modestly** in **complex real-world scenarios**, highlighting the importance of contextual evaluation.

Visualization Features (on which I have not worked but which helps presenting the project):

- Interactive maps for rainfall, river, and flood risk levels.
- Risk heatmaps for wet or significant flood-prone days.

Predicting Significant Wave Heights for Ocean Exploration

Context

This project, undertaken during an 8-hour long assessment, focuses on predicting significant wave heights, a critical aspect for ensuring the safety and efficiency of marine operations. The project utilizes data-driven methodologies to address challenges posed by the dynamic and complex nature of oceanic waves.

- **Objective**: To develop a predictive model for significant wave heights, aiding ocean exploration and operational planning.

Methodology

- Variable Analysis and encoding, along with reduction of dimensionality using **PCA**
- Cleaning and Transformation: Handling missing values, scaling numerical features, and ensuring consistency in units.

Model Development

- **Algorithm Selection:** A range of machine learning models (e.g., linear regression, decision trees, or gradient boosting) have been considered.
- **Validation and Tuning:** Hyperparameter tuning and cross-validation are employed to optimize model accuracy and generalizability. Permutation of features have also been used to verify which features were efficient.
- **Results and Evaluation:** Metrics such as RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) are used to assess model performance.
- **Comparison of baseline and advanced models** highlights the importance of feature engineering and non-linear methods in capturing wave dynamics.

Key Takeaways

- **Understanding Ocean Dynamics:** Wave heights are influenced by multiple interdependent factors, including meteorological and geographical variables.
- **Balancing Simplicity and Complexity:** While complex models can capture non-linear relationships, simpler models may provide interpretability, crucial for operational decision-making.
- **Future Directions:**
    - Incorporating real-time data streams (e.g., satellite and buoy data) for dynamic prediction updates.