

Predictive Analysis of Delhivery Logistics

Role: Team Lead & Machine Learning Engineer

Table of Contents

- Project Overview
- Key Responsibilities
- Technologies Used
- Key Achievements
- Additional Notes

1. Project Overview

Orchestrated an end-to-end predictive analytics project aimed at revolutionizing logistics efficiency by predicting delivery durations with unprecedented accuracy. Led a team of five in developing a scalable solution deployed in production, transforming business operations through actionable insights.

2. Key Responsibilities

Data Analysis and Preprocessing

- Built and optimized scalable data pipelines to process extensive datasets.
- Conducted advanced EDA to extract insights, identify anomalies, and inform feature engineering strategies.
- Utilized Python libraries (Pandas, NumPy, Matplotlib, Seaborn) for data visualization and exploration.

Model Development and Optimization

- Architected a Stacking Regressor achieving **MAE: 0.0021** and **R²: 0.9998**, outperforming industry standards.
- Evaluated and optimized machine learning algorithms including XGBoost, Gradient Boosting, and Random Forest.
- Automated **hyperparameter tuning** using GridSearchCV and RandomizedSearchCV for optimal performance.

Deployment and API Development

- Developed and deployed RESTful APIs using Django and Flask frameworks.
- Leveraged Docker for containerization and Kubernetes for scalable deployment.
- Integrated CI/CD pipelines via GitHub Actions to streamline deployment processes.

Leadership and Team Management

- Directed a team of 5 members, assigning tasks, conducting code reviews, and ensuring project milestones were met.
- Adopted Agile methodologies, utilizing daily scrums and sprint retrospectives for effective collaboration.

Documentation and Testing

- Authored detailed documentation outlining the data pipeline, model architecture, and deployment strategies.
- Implemented robust unit testing and A/B testing to ensure system reliability and scalability.

3. Technologies Used

- **Programming Languages:** Python (NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn)
- **Machine Learning:** XGBoost, Random Forest, Stacking Regressor, Gradient Boosting
- **Frameworks and Tools:** Flask, Django, Docker, Kubernetes, GitHub Actions
- **Database:** SQL
- **Version Control:** Git, GitHub

4. Key Achievements

- Reduced delivery duration prediction errors by 99%, enhancing logistics efficiency.
- Delivered a production-ready system, demonstrating real-world adoption by stakeholders.
- Exceeded performance benchmarks with a scalable and highly accurate predictive model.

5. Additional Notes

- **This project is part of a private repository forked from an organizational repository.**
- Detailed proof of work, including code snippets and technical contributions, is available **upon request.**
- This version organizes your work into sections for easy navigation while maintaining a high-tech, impactful tone. Let me know if you'd like to refine or add anything!