Project Overview File

InboxIQ–Intelligent Email Filtering with ML

Winter in Data Science (WiDS)

December 1, 2024

Welcome to InboxIQ!

Embark on an exciting journey into the world of Machine Learning and Natural Language Processing! As your mentors, we're thrilled to guide you through the InboxIQ project, where you'll develop an intelligent email classification system that can distinguish between ham and spam messages. This project isn't just about coding – it's an opportunity to dive deep into real-world applications of ML, enhance your problem-solving skills, and create something meaningful.

We've structured this project to ensure you gain both theoretical knowledge and practical experience while maintaining a comfortable learning pace. Regular catch-up meetings will help us track your progress and provide timely support. For project-related communications, resources, and file sharing, we'll primarily use Discord. Feel free to reach out to us on WhatsApp for any specific questions or concerns.

Project Timeline

Week 1

Familiarization with Python libraries: NumPy, Pandas, and Matplotlib for basic data analysis. A small programming assignment will be given on data analysis using the tools learned this week.

- **Day 1-2:** Access `Python_Examples_CS_DSA.zip` from our GitHub repository. These carefully curated Python programs will strengthen your foundation. Feel free to skip to the next section if you already have Python experience.
- **Day 3:** Master NumPy fundamentals through this comprehensive tutorial: NumPy Tutorial
- **Day 4:** Dive into Pandas and Matplotlib:
  - Pandas: Pandas Tutorial
  - Matplotlib: Matplotlib Tutorial
- **Day 5-7:** Complete a hands-on assignment focusing on CSV manipulation with NumPy and data visualization tasks.

You'll receive both the CSV file and assignment details on Day 1. If you're already familiar with the concepts, feel free to complete the assignment early and move on to Week 2 tasks. Otherwise, we recommend following the suggested timeline. Assignment-1 must be submitted by the end of Week 2.

Weeks 2 and 3

Introduction to ML fundamentals and Deep Learning concepts. We'll provide structured learning materials time to time and two practical assignments to reinforce your understanding.

Essential Resources:

- StatQuest ML Fundamentals: Series contains concepts explained in small videos
- Stanford CS229: For more mathematically oriented ML concepts
- CampusX ML: Everything simplified (optional, Hindi)
- Kaggle Tutorials:
  - Intro to Deep Learning with TensorFlow basics
  - RNN Tutorial with hands-on implementation

Note: Don't worry about watching entire playlists – we'll share specific, limited video recommendations weekly.

Weeks 4 and 5

These crucial weeks are dedicated to the main project implementation. You'll apply everything you've learned to build an intelligent email classification system. Here's what to expect:

- **Dataset Introduction:** You'll receive the project dataset at the beginning of Week 3. Once you've completed Assignments 2 and 3, you can start working on the final project immediately.
- **Project Development:**
  - Data preprocessing and exploration
  - Feature building and selection
  - Model development and training
  - Performance optimization and validation

For initial validation, use an 80-20 split of the provided dataset (80% training, 20% validation). We'll provide a separate test dataset by the end of Week 4 for final evaluation.

- **Regular Check-ins:** Mandatory twice-weekly meetings to:
  - Review your progress and address challenges
  - Discuss your approach and potential improvements
  - Keep you on track for successful completion

Final Submission

Your project submission should be completed by three days after Week 5 and should include:

- A well-structured Jupyter notebook with:
  - Clear markdown documentation
  - Commented code sections
  - Visualization of results

- A comprehensive project report covering:
  - Methodology and approach
  - Challenges faced and solutions implemented
  - Key learnings and insights
  - Future improvement possibilities

Your Mentors:
Feel free to reach out to us with any questions or concerns:

- **Yaswanth:** 7989068522
- **Avyay:** 9550104004

We're excited to guide you through this learning journey!