Date: 2024-11-29

DSMP Projects

Project 1

1. **Project Statement:** This project aims to create an interactive dashboard that displays North American oil and gas production data. The dashboard will allow users to visualize production trends across different regions, using maps and filters. It will integrate data from government sources (EIA, CER) and use machine learning to predict future trends, helping various teams to analyze and forecast production more efficiently.
2. **Input Dataset:** EIA (oil, gas), CER (oil, gas) - there's a way to retrieve this data through API or using web scraping libraries.
3. **Instructions for Mentees:**

Steps:

1. End-Users
    - Determine who the dashboard is for. This could include management teams, investment analysts, engineers, or other stakeholders in the oil and gas industry.
2. Functionality
    - Clarify the main features of the dashboard. What specific data or insights should it provide to users? For example, displaying oil and gas production over time, forecasting trends, and comparing production across regions.
3. Design
    - Create a rough design or wireframe for the dashboard layout. Decide on key elements like maps, graphs, and filters. Think about how users will interact with the data and what functionalities are needed (for example, toggle for oil price, region selection, etc.).
4. Data retrieval
    - Use requests or urllib to pull data from government APIs.
    - For web scraping, use BeautifulSoup or Scrapy to extract information from websites.
5. Data analysis and cleaning
    - Clean and organize data with pandas and numpy.

6. Machine learning forecasting
    - Use scikit-learn for machine learning models.
7. Interactive map visualization
    - Create interactive maps with folium, plotly, and geopandas.
8. Wrapping as a service
    - Build the web interface using Streamlit or Flask.
    - Deploy the project using Docker for containerization and services like Heroku, AWS, or Google Cloud.

Contact Information: Rushad Rakhimov. Email: **rrakhimov@mcdan.com**

Project2: Volve Bottomhole Pressure Prediction

Project Statement

Production data is ubiquitious in the oil and gas industry. Fluid and pressure are linked by the laws of physics, and understanding this relationship is key to ensuring production. In this project, we will use production data from the Volve field to build a model that can predict bottomhole pressure.

Read more about the Volve field and dataset here: https://www.equinor.com/energy/volve-data-sharing

Input Dataset

The input dataset can be downloaded from the volve data sharing portal. But to save everyone some time, I've already downloaded it and put it in github (https://github.com/ScottHMcKean/volve_bhp). An example of a similar analysis can be found here: https://github.com/KarimMAJDOUB/Volve-field-machine-learning/tree/main/BHP%20Prediction.

Initial Instructions

Your mission, should you choose to accept it, is to build a model that can predict bottomhole pressure. That will likely involve the following milestones:

1. Parse the data

2. Do some exploratory data analysis and produce a report

3. Do data cleaning and feature engineering - how do you deal with outliers, errors, and missing data?

4. Build a model to predict bottomhole pressure

Some questions you might want to consider:

- How do you deal with the small amount of data?

- Are you going to build a global or local model?

- How are you going to do cross validation?

- How are you doing to normalize the data?

- What kind of models are applicable?

- Have you considered the law of physics?

Contact Information: Contact Scott McKean at scott.mckean@databricks.com.

Project 3: Volve Drilling Anomaly Detection

Project Statement

Real-time information is critical to drilling operations. The ability to detect anomalies in the drilling process is key to ensuring safety and efficiency. In this project, we will use drilling data from the Volve field to build a model that can predict anomalies in the drilling process.

Read more about the Volve field and dataset here: https://www.equinor.com/energy/volve-data-sharing

Input Dataset

Real time drilling data is sent in XML files with a WITSML schema. Processing these files is a little brutal and a data engineering project in and of itself. Fortunately, Stavanger University has already done a lot of the heavy lifting for us and published a processed version of the data here: https://www.ux.uis.no/~atunkiel/file_list.html.

Initial Instructions

Your mission, should you choose to accept it, is to build a model that can predict anomalies in the drilling process. That will likely involve the following milestones:

1. Download and and parse the data

2. Do some exploratory data analysis and product a report

3. Distinguish what is an error in the data and what is anomalous drilling behaviour

4. With an idea of what is anomalous, can you build a model to inform the drilling crew what is anomalous?

Some questions you might want to consider:

- How will this model be used in the field?

- What kind of back testing and validation is appropriate for a model like this?

- Are you build a global or local model (one model to rule them all, or one model per well)?

- How performant is your model? How did you prove this out?

- What causes the anomalies?

Contact Information: Contact Scott McKean at scott.mckean@databricks.com.