

Pooja Mishra

+91-79920 82636 | ✉ mpooja25@myyahoo.com

🌐 <https://github.com/Vijaylekha25> 🌐 <https://www.linkedin.com/in/pooja-mishra5a0630269/>

Sant Kabir Nagar, Uttar Pradesh

Summary

Highly motivated and detail-oriented recent Bachelor of Technology graduate with a strong foundation in data analysis, seeking a Data Analyst, Junior Data Analyst position. Proficient in **Python**, **Machine learning**, **Deep learning**, **SQL**, **Power BI**, and **Advance Excel**, with a passion for leveraging data to drive business insights and decisions.

Work Experience

Junior ML Engineer (Internship)

Omdena (1 months, July 2024- Present)

Project:

CanPolicy Insight - Clear Insights into Canadian Policies Leveraging LLM:

(<https://www.omdena.com/chapter-challenges/canpolicy-insight-clear-insights-into-canadian-policies-leveraging-llm>)

- Contribute to the data collection by sourcing and scraping records of Canadian parliamentary debates, including Hansard transcripts and committee reports using **Python**.
- Process the collected data to ensure accuracy and usability using Python (**pandas** and **NumPy**) and collected the articles for further analysis in **Excel sheet**.
- Analyze the processed data to identify key themes, topics, and impacts of the discussions using Python (pandas) and parsing them to analyze it.
- Apply **NLP** and **Machine Learning** techniques (such as **text summarization** and **topic modeling**) in Python (using libraries like **NLTK**, **SpaCy**, and **scikit-learn**) to simplify complex parliamentary language into concise summaries.
- Develop an API in Python (using **Flask** or **FastAPI**) that provides simplified summaries and insights based on user queries.
- Create and host a user-friendly platform using **Python (Flask or Django)** where citizens can access simplified parliamentary summaries and understand policy impacts.
- Create the interactive dashboard using **Power BI** for Collaborators.

Projects (Personal):

Diwali Sales Analysis and sales predictions

- Utilized advanced Excel techniques for further data exploration and analysis, including creating **pivot tables** and **charts** to **visualize** key **sales metrics**.
- Visualized every feature against the response variable and extracted insights using **advanced Excel charts**.
- Applied conditional formatting to highlight **key performance indicators (KPIs)** and outliers in the sales data.
- **Queried SQL** to extract and aggregate sales data, including **filtering**, **grouping**, and **sorting** to analyze sales trends and patterns.

- Used Excel's **VLOOKUP** and **HLOOKUP** functions to merge datasets and enhance data integrity.
- Leveraged **Excel's data validation** and dropdown lists to standardize data entry and reduce errors.
- Created dynamic dashboards using **Excel's slicers** and **pivot charts** to provide interactive data views.
- Conducted data preprocessing using **Python** with **pandas**, including importing essential libraries (**pandas, seaborn, matplotlib**) and loading data from a **CSV file**.
- Explored the data by checking dimensions, displaying initial rows, and using **df.info** to understand data types and missing values.
- Cleaned the data by identifying and removing irrelevant columns (e.g., 'Status', 'unnamed1'), handling missing values, converting the 'Amount' column to integer format, and optionally renaming columns for improved readability.
- Described the data using **df.describe** to calculate **summary statistics** like **mean, standard deviation, minimum, and maximum** values for numerical columns.
- Conducted the Exploratory data Analysis (**EDA**) to find the insights in existing data.
- Conducted **Data Visualizations** using **Python** with **time series plots** to study price variation patterns of commodities using financial and economic indicators, employing **Plotly, Seaborn, and Matplotlib**.
- Converted the categorical variables into numerical using **OneHotEncoder** library in **Sci-Kit** learn.
- Split the dataset into train and test using train test library from Sci-kit library.
- Built various supervised machine learning models like **linear regression, Random Forest Regressor, XGBoost Regressor**, and some other as well using pipeline library from **Sklearn**.
- Created **pickle** files of the models using the **pickle** library in **Python** and developed a web application using **Flask** with the help of frontend team.
- Deployed the project on **AWS Elastic Beanstalk** after modeling using **AWS CodePipeline**, ensuring continuous integration and deployment.

SKILLS:

- **Programming Languages:** Python, SQL
- **Data Analysis & Visualization:** Advanced Excel, Tableau, Power BI, Matplotlib, Seaborn, Plotly.
- **Data Science/Machine Learning/Deep Learning:** EDA, Supervised Learning Algorithms, Unsupervised Learning Algorithms, Feature Engineering, Feature Selection and Extraction, Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Natural Language Processing (NLP), Time Series Analysis, Model Evaluation and Hyperparameter Tuning.
- **Tools & Technologies:** Git, AWS, Flask, ElasticBeanstalk

Academic Qualifications:

Bachelor Of Technology from Uttar Pradesh Technical University, Lucknow, India (2021-2024)

Diploma in (Plastic Mould Technology) from Government Polytechnic Gorakhpur, U.P (2017 – 2020)

Certifications:

Master's in data science from PW Skills / iNeuron (2023-2024)

(<https://ineuron.ai/> , <https://pwwskills.com/blog/foundations-of-data-science/>)