APOORV PATHAK

DATA ANALYST INTERN

CONTACT

(+91)7415855937

apoorvgunjanpathak@gmail.com

Raj Labdhi Heritage, Gandhinagar
Gujarat, India 382421

LINKS

Medium **@apoorvnpathak**

LinkedIn **Apoorv Pathak**

GitHub **ApoorvPathak2003**

Stack Overflow **Apoorv Gunjan Pathak**

PROFILE

I am a **Computer Science student** pursuing **B.Tech(3rd year)** at **IIIT Vadodara**. Throughout my academic career, I have developed a strong interest in **Data Science and Analytics**, **Machine Learning**, and **Deep Learning** domain. I also have a basic understanding of **economics** and **finance** and am ready to learn and explore a new area of work.

I have **experience of 3 months** as a **data science intern** at **INNODATATICS** where I learned to build **pipeline architecture**. I have also worked on several group projects in which I completed **Data Analysis** and used **Data Visualization** tools, **Machine Learning**, and **Deep Neural Network** techniques to accomplish my task. Along with this, I have experience working as a freelancer in which I gained hands-on experience in **Technical Documentation** thesis work, and also worked on multiple data science side projects.

I admire the company's **commitment** to its work. As someone **passionate** about similar areas of work and ethics, I believe that I would thrive in an environment where I can make a meaningful impact. I am also drawn to the company's culture, which aligns with my values. As a **highly motivated**, **detail-oriented**, and **analytical decision-maker**, I am extremely excited to apply for the data science & analyst intern position in your company. With the proven ability to **lead cross-functional teams** and **deliver results** in **fast-paced environments**. I am eager to hit the ground running and contribute to the team's success from day one.

SKILLS

Python	★★★★★	VS Code	★★★★
MS Power BI	★★★★	Figma/Canva	★★★
PostgreSQL	★★★★	Java	★★★
SQL	★★★★	LaTeX	★★★
MS Powerpoint	★★★★	Streamlit/Django	★★
MS Excel	★★★★	Git	★★

APOORV PATHAK

DATA ANALYST INTERN

PERSONAL PROJECTS

Project 1: Market Basket Analysis

GitHub: Project Link

- Market Basket Analysis is a data mining technique used quite frequently by the shop-owners and retailers to understand the shopping behavior and the purchasing pattern of their customers. This project contains data analysis and data visualization which helps understand the customers' purchase behavior and also analyzes the number of accidents in different cities corresponding to the sales of liquor in that state.
- **Problem Statement:** To analyze and provide insights into the purchasing behavior of customers and to help retailers make decisions about their sales strategies to increase profits.
- Initially, we collected the data from internet resources. Our project primarily utilizes 3 datasets i.e., monthly sales data, grocery description dataset, and items bought dataset.
- Metrics and algorithms used to compute the details of the customer behavior are **Antecedent**, **Consequent**, **Support**, **Confidence**, **Lift**, **Zhang's Metric**, **Centrality**, **Silhouette Coefficient**, and **Apriori Algorithm**.
- In the further step, **data visualization** was done using **matplotlib**, and **seaborn** library; in this step, **customer segmentation**, **heatmap plot**, **item's support matrix**, **basic statistical charts**, **sales data visualization**, and **centrality network plots** were drawn.
- Detailed code and analysis can be obtained on my GitHub account, link at the top-right corner.

Project 2: Jarvis

GitHub: Project Link

- Jarvis is an audio assistant that helps the user to shift through the tasks effectively and perform multiple activities simultaneously.
- **Functionalities** associated with Jarvis:
  - Sending emails.
  - Searching Wikipedia.
  - Opening YouTube & Google.
  - Playing local music.
  - Opening VS code.
  - Google search.
  - Tells the current time.
  - Greets the master as per time-zone and on their birthday
- Detailed code can be obtained on my GitHub account, link at the top-right corner.

Project 3: Address Book

GitHub: Project Link

Project 4: Chat Room

GitHub: Project Link

Project 5: Tic-Tac-Toe

GitHub: Project Link

Project 6: Business Sales Analytics(Power BI)

GitHub: Project Link

APOORV PATHAK

DATA ANALYST INTERN

WORK EXPERIENCE(INDUSTRIAL)

DATA SCIENCE INTERN - INNODATATICS Jan 2023 - Mar 2023

Project 1: Anti-Money Laundering System

- This project was built to warn the system in case any fraudulent transaction occurs; the system would detect any fraudulent transaction based on the machine-learning model deployed. In this project, we strictly followed **CRISP-ML(Q)** technique to complete the project. My task was to give a complete overview of this system.
- In the first step, I was supposed to collect data from multiple open-source platforms and store them on a local device. Further, I cleaned, pre-processed, and analyzed this collected data to gain insights from the data; this task was accomplished with the help of **pandas**, **numpy**, **matplotlib**, **seaborn**, **scikit-learn**, **scipy**, and **joblib** library.
- After completing the analysis, the next step I was asked to do was build a pipeline architecture to automate the machine learning workflow by transforming the data that can be analyzed to achieve the output.
- After completing the analysis of the complete dataset, I was given the responsibility to build multiple machine learning models like **XGBoost**, **Random Forest Classifier**, **LightBGM**, **Logistic Regression**, and **Decision Tree** models. Out of the above-mentioned models, **LightBGM** was the **best-performing** model with an **accuracy** of **almost 99.3%**.
- Later on, I was also supposed to build a creative dashboard that would reflect the insights drawn from the analyzed dataset using the **MS Power BI** tool.
- For the successful deployment of the model, we used **streamlit** and **Django** frameworks.

Project 2: NLP Project - Voice Modulation System

- This project was built to recognize the emotions of the speakers using audio as input.
- I collected some audio files from YouTube and extracted audio from them. I was also given a dataset by the team to complete this analysis.
- In the first step, I explored the dataset like **statistical properties**, **wave plots**, and **spectrogram** visualization. In the next step, the noise(**time stretching**, **pitch shifting**, **time shifting**) was introduced to let the model learn on a more generalized dataset.
- In the next step, we begin extracting features of the audio file using the **librosa** library; these extracted features were then stored in the data frame and saved in a CSV file for performing further computation.
- In the next step, we performed data preparation where **One-Hot Encoding** and **Standard Scaling** technique was used; this converts the data as per the model building requirement.
- After completing all the above-mentioned phases, we build multiple machine-learning models like **MLP Classifier**, **Decision Tree Classifier**, **Random Forest Classifier**, and **LightBGM** model and a few deep-learning neural networks like **LSTM**, and **CNN**. We also hyperparameter-tuned these models to get the best combination of the parameters.
- **LightBGM** model was the **best-performing model** with an **accuracy** of almost **71%**.
- We also used an **AutoML library TPOT** to complete this analysis and gain insights from the pipeline and model obtained from it.

APOORV PATHAK

DATA ANALYST INTERN

WORK EXPERIENCE

Freelancer — April 2023 - Contd.

- While working as a freelancer, I have worked with clients from Ireland, Canada, and India and helped them deliver **technical documentation** and **coding** related work for data science projects.
- I am proficient in using tools like **MS Word**, **MS Excel**, **Google Sheets**, and **Google Docs**.
- In this period, I have worked on and completed projects for almost **10 clients**. I am **detail oriented** and a **collaborative team player**, delivering work to my clients with a **satisfactory ratio** of **over 95%**.

ROLE OF RESPONSIBILTY

Secretary

The Pensieve Club, Literature Committee

Jun 2021 - ***Present***

- Official Member of the **Core Support Team** of the **TEDxIIITV** edition **(Speaker - CA Nandini Agarwal and Anuj Kumar Sharma)**.
- Experience in working with teams and organizing multiple events and competitions in the institute.

Active Member

The DOT Club, Designing Committee

Dec 2021 - ***Present***

- Designed a couple of poster for college events and competitions.

Active Member

The Obscura Club, Photography Committee

Feb 2022 - ***Present***

ADDITIONAL SKILLS

- *Strategic Thinking*
- *Creative Problem-Solving*
- *Article Writing*
- *Social Media Management*
- *Data Analysis*
- *Project Management*
- *Leadership*
- *Team Management*
- *Communication*
- *Collaboration*
- *Business Analytics*

HOBBIES & INTERESTS

- Book Reading
- Economics & Finance
- Article & Blog Writing
- Gymming
- Event Management
- Cooking