

Deepak Kumar Jha

Data Scientist

Pune, India

Ph: +91 9943898886

Github: <https://github.com/LEGEND1099>

deepakkumarjha178531@gmail.com

Profile

I currently work at one of the largest NBFCs in the country, where I contribute to business growth through data analysis. I excel in collecting and monitoring data from various sources, including DataLake, SQL, and Excel.

Education

- B. Tech in Information Technology, VIT Vellore, CGPA: 8.52/10
- CBSE, 12th Boards, FIITJEE, 72%
- CBSE 10th, Navy Children School, CGPA 10/10

Jul 2018 - Jul 2022
Apr 2016 - Apr 2017
Apr 2014 - Apr 2015

Employment History

● Unit Manager - Data Science BFL App Team

Jul 2022 - Present
Pune, India

- Constructed a K-Means Clustering model to categorize customers into clusters and recommended search keywords tailored to each cluster based on their activity within the app. This initiative enhanced the role of search as an integral component of the customer journey within the App and helped grow the search traffic by 82% in 6 months.
- Conducted customer cohort analysis to facilitate the scaling of the Insurance Business. This involved constructing propensity models to identify users more likely to engage with the App's offerings. Additionally, I contributed to enhancing offer pools for sub-businesses within the Insurance division, playing a significant role in achieving a remarkable monthly sales growth of over 200%.
- Engineered automated triggers to enable businesses to target customers in immediate need of loans and helped identify a pool of over 5Lakh high value customers by leveraging various customer data sources, including SMS data. I applied RegEx for data mining.

● Intern - Data Science BFL App Team

Jan 2022 - Jul 2022
Pune, India

- Played a pivotal role in enhancing the search feature within the mobile application, which was in its nascent stage when I joined the project. I uncovered valuable trends and usage patterns among our user base of over 30 million customers. These discoveries were instrumental in improving the search product by introducing advanced features, such as personalized recommendations and optimizing CTA elements within the search interface.
- Made substantial contributions to the Insurance team's app journey redesign, offering insights into user behavior and creating effective PowerBI dashboards.

● Backend Engineer - CreatorHQ

Jun 2021 - Dec 2021
Remote, India

- The team worked on a platform enabling peer-to-peer donations and charity through the Ethereum blockchain. My role focused on backend development, where I built their data storage platform using Railway, a cloud data storage service.

● Data Science - C4 Projects

Feb 2021 - Dec 2021
Vellore, India

- Assisted the team in analyzing EdTech trends by scraping X (formerly Twitter) with BeautifulSoup and AWS Linux Servers. Leveraged LLMs like Pegasus for data mining and summarization, contributing to Investor Presentation decks that secured our first-round investments.
- Served as a mentor for Data Science, instructing the inaugural group of students on our platform, by taking a [ML Crash Course](#).

Projects

● [Phishing Website Detection \(PhishProtect\)](#)

Jan 2021 - Mar 2021

- Developed a full-stack web application for identifying phishing websites. We evaluated various ML algorithms and neural networks, achieving a 93% accuracy rate. The application was built using the Streamlit Framework for web app and hosting, and we implemented low-level multi-threading to improve response times.
- Technologies Used: Python, Random Forest Classifiers, KNN, Neural Networks and Streamlit.

● [Burn Rate Predictor/Happy Hours \(SAP Hackathon\)](#)

Mar 2021 - Apr 2021

- We created a platform within a project presented at SAP Hack 2021 to predict employee burnout in organizations. We utilized a Logistic Regression Model for predicting burnout probability.
- Technologies Used: Python, Logistic Regression, Decision Trees, Random Forest.

Technical Skills

Programming Languages & tools - Python,PySpark, SQL,JupyterNotebook, Azure Databricks

ML Algorithms - KNN, Logistic Regression, Decision trees, Random Forest, CNNs