

# Hamza Imtiyaz Ghojaria

Email: [hamza.ghojaria123@gmail.com](mailto:hamza.ghojaria123@gmail.com) | LinkedIn: [hamzaghojaria](#) | DOB: 29th October 1996 | +91 9892203101 (IN)

## Education:

- **Mumbai University: Sardar Patel Institute Of Technology ( May 2020 - May 2022 )**  
Masters of Technology in Computer Engineering ( **CGPA 3.2 / 4** )

## Skills:

**Data Science:** Python, Generative AI, Machine Learning, SQL, REST API, Flask, Fast API, Analytics, Deep Learning, NLP.

**Software:** Visual Studio Code, Excel, Jupyter Notebook, Pycharm, Anaconda, Spyder, MSSQL, Tableau, PowerBi.

**Cloud:** Azure (Containers, Blob, Document Intelligence), GCP

## Projects:

- **Document Compare Engine:** The objective of this project is to extract **semantically meaningful sentences** from user-uploaded documents. The engine enables users to compare source and query files across various formats, including PDF, Word, JSON, and Excel. Structured formats like JSON and Excel undergo pre-processing to create a unified format, while PDF and Word documents are processed using **Azure AI Document Intelligence**, converting them into JSON for flattening and standardization. Based on user selection, the unified data is filtered, chunked, embedded, and stored in a Vector Database. Retrieval is facilitated through prompts with **Azure Search**. The entire process is managed through a **FastAPI for the backend** and **ReactJs for the front end**, leading to a **60% reduction in errors** and an estimated 70% savings in time previously spent on manual document processing.
- **Recommendation Engine-** Designed a recommendation engine to increase user retention on the website. The two models that were implemented were **popularity** and **sequence models**. As per the business problem, the sequence-based model is implemented based on the **FP growth algorithm**. For the popularity-based model, the popular product is based on the **buying frequency** in each Pincode /State /District. This resulted in a **50% increase** in user retention time and a **10% boost** in product sales.
- **Referral Model:** Designed and implemented a propensity-based machine learning model utilizing the **XGBoost Algorithm**. The model is capable of accurately predicting conversion likelihood for each prospect. Through iterative refinement and validation. The model contributed to a **25% increase** in the **conversion rate** within a short timeframe.
- **Multiple ML Model Implementation-** Developed an end-to-end machine learning pipeline to conduct a comparative analysis of multiple machine learning model implementations simultaneously. Users can select a dataset, where missing values are handled. Followed by data visualization using the **Pandas Profiling library**. The pipeline presents a list of many models from **Regression, Classification, and Clustering**. upon selecting a category, all models are executed concurrently, providing accuracy metrics for each. The backend is built using **Python** and **Flask**, while the frontend is developed with **HTML, CSS, and JavaScript**. This approach greatly reduces the team's efforts by **60%** when selecting the optimal machine learning model that delivers the **best results**.

## Professional Experience:

### **LTIMindtree - Data Scientist: ( Mar 2024 - Present )**

- The Python-based document comparison tool, designed for a finance client, effectively automated the identification of discrepancies in financial documents, ensuring accuracy and consistency in reporting.
- The tool, which was developed using Python, efficiently streamlined the review process by generating detailed reports on differences, thereby aiding compliance efforts.
- The Python-powered document comparison engine, not only automated the identification of discrepancies but also provided valuable insights into the inconsistencies, enabling stakeholders to make informed decisions.

### **Hansa Cequity - Associate Data Science and AI: ( May 2022 - Mar 2024 )**

- Implemented a recommendation engine using Python for an e-commerce website by utilizing machine learning techniques to provide personalized product recommendations.
- Designed propensity-based models using machine learning models (**Logistic Regression, Random Forest, XGBoost, Decision Tree**) to examine large datasets and provide actionable insights for retail and e-commerce clients.
- Developed transaction data analysis tools to assess impact analysis and simulate business performance, helping clients optimize revenue and minimize costs.

### **CRISIL - Data Science Intern: ( Dec 2021 - Apr 2022 )**

- Developed an end-to-end machine learning pipeline incorporating multiple Machine learning models.
- Implemented **Flask** using **REST API** to interact with the client-side application and present results.
- Implemented scalable architecture using **Python** for an **Accounting firm** based on keywords like Debt, Equity, Revenue, Cost, etc. to be **scraped** from the required documents and keep a check on optimal results.
- Worked on a business problem where a table needed to be **scraped** from the pdf's set of pages and used **LabelImg** to annotate those pages & pass the optimized outcomes for further development of the product.
- Designed and coded in Python to solve business problems, **data manipulation & data aggregation** for clients and teams.

### **Media.net - Associate Software Test Engineer: ( Mar 2019 - Mar 2021 )**

- Proficient in designing and maintaining **Python** scripts to automate test cases and template codes, leading to increased efficiency and reduced manual effort in testing and development.
- Implementing **Web Scraping** methods to extract desired data and keep a check on parameters implemented by the development team.
- Demonstrated expertise in Python and its associated libraries and frameworks, resulting in improved product quality and reduced

time-to-market for clients.

**AlphaCentrix - Team Member QA & Support: ( Jun 2018 - Feb 2019 )**

- Maintained and optimized **ETL tool using VB Script**, ensuring smooth operations through regular updates and code modifications.
- Executed data migration processes for the updated ETL tool, ensuring accurate data transfer and minimal client downtime.
- Managed multiple finance clients' operations, loading trades, equity, and bonds to the system while maintaining **SQL query updates** based on client requests.

**Certifications:**

Introduction to Generative AI Course: Google, Google Cloud Certified Professional Cloud Architect Course: Qwiklabs, IBM WatsonX Foundations Course: IBM, Introduction to Large Language Models Course: Google, Develop Generative AI solutions with Azure OpenAI Service: Microsoft, Text Prompt Engineering Techniques: Google, LangChain for LLM Application Development: Deep Learning.ai.