

Gourav Suman

Mobile: +91-9122014763 | Email: [gourav.suman.js@gmail.com](mailto:gourav.suman.js@gmail.com) | <https://github.com/GOURAVSUMAN>

LinkedIn Profile: <https://www.linkedin.com/in/gourav-suman-data-scientist/>

HackerRank 5 Star Badge (Python and SQL): @gourav\_bjn19bdv1

---

**I graduated from SP JAIN in Data Science & Engineering Program Mumbai campus with 3.5+ years of experience into Data Science and Analytics am looking for a full-time position in a reputed company where I can expand my learnings and develop products, mentor and guide team based on the experience and learning**

**Skills:** Python, NumPy and Pandas, SQL, Machine Learning, Deep Learning (NLP, NER, Grammar based metadata and much more.), Hadoop, SPSS MODELER, Tableau, RASA Chat Bot, AWS Lambda, AWS EC2, AWS S3, AWS Kendra and other AWS module, Git, Search Engen Bot, Docker Container and Image building, Flask API, FAST API, Chatbot, NLP base flowchart, Web developments, web apps, Python UI webapp frameworks like Striemlit, Anvil, pynecone etc

### Education

- 2019-2020 | Post Graduate in Data Science & Engineering Program (Full time campus) | SP Jain (DASCA Certified), Mumbai IN (GPA 2.9/4.0)
- 2015-2018 | B.Com. (Hons), Calcutta University (53%)

### Certification

- Certification in Prescriptive Analytics
- Financial Analytics (Banking and Insurance)
- IBM Certificate Python for Data Science
- Certification in Google Analytics for beginners
- Specialization in NLP, Chatbots, Rule based engine, Search engine/bot, ML modeling's etc.

### Experience

#### Data Scientist | Liveage Technologies | Since Aug 2021

- Analyse, monitor and develop analytical models to estimate annual, monthly, daily platform returns and other key metrics and weekly tracking of AOP vs actual returns performance. Monitor metrics across organization for all departments. Work with BI teams to maintain dashboards across organization
- Owning the full Power BI lifecycle from requirements gathering through design & development, and through release phases. Articulate key process flows, manage inputs and priorities. Execute quantitative analysis that translates data into actionable insights. Design and Support test and deployment of new products.
- Plan, analyse, monitor and direct the gathering of data, assessing data validity and synthesizing data into large analytics datasets to support project goals. Utilize big data analytics and advanced data science techniques to identify trends and discrepancies in data. Use deep learning & image recognition project goals.
- Influence various cross functional teams/stakeholders within the organization to meet goals & planning timelines. Describe data sets, sources, and structures. Analyse and develop reports to make data more meaningful to business initiatives, analyze data quantitatively / qualitatively and communicate findings.

### Corporate Projects

- Developed NLP based FAQ generation model and API. (Project Leader- Manager): Developed end to end Question/Answer/Summary generation from English context using NLP techniques and different transformers-based models and deployed same using docker containers in AWS EC2 server
- Developed NLP based chatbot for the consumer portal. (Project Leader-Manager): Developed end to end Chatbot to handle user question for customer portal using NLP techniques and logic to make conversational level bot without 3<sup>rd</sup> party software or libraries and deployed same using docker containers in AWS EC2 server
- NLP based Flow chart information documents for legal and financial documents. (Project Leader- Manager): The product development includes the entire end-to-end product development and data extractions from the bunch of pdfs and structuring them into the consumable form as a part of Data Engineering and the product data pipeline. The technologies involved in the process are Python, Numpy, Pandas, AWS, NLP models like spacy models, dispy, punctuator pretrained models, Bert and Distil-Bert pretrained models etc

- Developed Chatbot using RASA as an SAS product: Developed end to end pipeline to create and deploy chatbot using RASA on single click for different clients just by sampling uploading different data frames/excels to handle custom customer requirements and expose ready to use API to end user.
- Developed Rule based DL models to qualify loan-based customer data. (Team activity): Developed rule-based loan qualification model using Python, SQL, AWS services. This involves extracting text data from PDF's first then extracting features and metadata from every sentence and heading level text and finally developing rule from sentences and store data in the form of SQL data tables.
- Developed UI Frontend app for product launch internally to be used by BA/Lawyers: Developed and designed simple UI that communicates with Databases to update and edit set of rules with the help of Python based web development frameworks and explored more than 3 such frameworks. The application is further deployed on EC2 and been exposed to be used by BA and Lawyers.

#### **Data Scientist | Infobiz LTD, London UK | July 2020 to July 2021**

- Analyse and build statistical models and machine learning algorithms. Apply semantic, natural language processing & understanding expertise. Productionize models and make those available at scale. Enhance data collection procedures to include info that is relevant for building analytic systems.
- Select features and optimizing classifiers using Machine Learning techniques. Develop automated processes and tools to monitor and analyze model performance and data accuracy. Leverage big data to discover patterns and solve strategic business problems using structured and unstructured data sets
- Design, follow and document best practices and processes. Assess the effectiveness and accuracy of new data sources and data gathering techniques. Ensures that proper and adequate model documentation and model risk mitigation processes are in place to appropriately assess and mitigate model risk.

#### **Corporate Projects**

- Developed Email Classification model using Bert NLP model: The model helps us to build email classification model using Bert NLP model, various other Deep Learning models and Python. The algorithms used are NLP Bert model, Multinomial Naive Bayes classifier, Passive Aggressive Classifier etc.
- Developed customer persona classifier for real estate clients using Machine Learning, Python Programming etc: The model helps us to predict the most promising clients who will be likely to buy property. The model was prepared using Machine learning algorithms like Logistic regression, Decision Tree, Random Forest, MLP Classifier, Naïve Bayes Classifier etc. using python programming languages.
- Developed Dashboard for higher management reporting using Dash and Plotly using Python Programming. This project includes the fundamental application of Python, HTML, and Dash.
- Developed Chatbot using python programming and integrated FAQ model with RASA bot. Developed chatbot using RASA module with the help of python program and integrated Deep Learning FAQ model for the chatbot using haystack python package using existing Q&A data which was frequently asked by the clients from the company. The performance of modern Question Answering Models (DISTILBERT) has seen improvements within the last year enabling many new opportunities for accessing information efficiently.

#### **Data Scientist | Edelweiss Financial Services (Mumbai) | Feb2020 to Jul 2020**

- Developed 3 different Predictive and Recommendation model for the Life Insurance, Medical Insurance and Travel Insurance Advisory team. The model predicts as well as recommend simultaneously to help their Advisory team to advice their clients based on their profile that which Life Insurance, Health Insurance, Travel Insurance product one must purchase and suggesting the best product. Model was built using Python Programming. The project involved, EDA, data Preparation and cleaning and applying multiple classification techniques which was further evaluated using AUC curve, Accuracy scores etc.
- Developed product classification model for different type of customer for their insurance products. The model developed my me helps the organization to classify the customer and further predict the product a customer is likely to buy and also appropriate products which a customer should buy so that it can fulfil their investment requirements. Prepared using Machine learning algorithms like Logistic regression, Decision Tree, Random Forest, MLP Classifier, Naïve BayesClassifier etc. using python programming languages.

**Gourav Suman**