



# Rishav Saigal

Data Scientist

📍 J/I - 4, Aswini Nagar, Baguiati, Kolkata, West Bengal, 700159, India

📅 09/02/1996 ✉️ [rishavsaigal@gmail.com](mailto:rishavsaigal@gmail.com) 📞 9804735299

🌐 [rishav1996](#)

in [rishav-s-4bb50311a](#)

With 5+ years of expertise in AI and ML technology development, I have a strong track record in creating large-scale AI/ML projects and fine-tuning hyperparameters. My skills include model training, building AI/ML solutions, and an ability to quickly learn and observe. I am curious about emerging technologies and have experience in constructing MLOPS pipelines.

## WORK EXPERIENCE

**Genpact** (08/06/2022 - 03/11/2023)  
Assistant Manager

- Currently focused on Time Series MLOps Pipelines and Data Robot.
- Successfully managed multiple parallel Proof of Concepts (POCs) including Python DASH Visualization and Real-Time Speech Analytics using AWS Transcribe & Comprehend in conjunction with Hugging Face.
- Developed a TSP solution POC tailored for diverse sales representatives.
- Conducted customer demonstrations of internally developed solutions and gathered business requirements for various use cases.
- Led a team of six resources, providing effort estimates and overseeing architectural adjustments to deliver scalable solutions for clients.
- Proficiently implemented AI/ML solutions utilizing Lambda Functions on the AWS platform.

**Capgemini** (11/04/2018 - 07/06/2022)  
Consultant

- Proficient in working with various Time Series Forecasting models such as ARIMA, Croston, ETS (Exponential Smoothing), LSTM, RNN, and TBATs.
- Possess a strong knowledge of Machine Learning techniques including Random Forest, Decision Trees, Linear Regression, Logistic Regression, XGBoost, and more.
- Skilled in designing and constructing Deep Learning models, with expertise in Keras and TensorFlow, including different layers like CNN, Dense, Dropout, and others.
- Familiarity with Survival analysis, particularly well-versed in cox proportional hazard and Cox time varying models.
- Engaged with approximately 8-10 clients, conducting client demos of internally developed solutions and gathering business requirements for use cases.
- Successfully led a team of 15 resources, handling the cost and effort estimation of solutions, and implementing architecture level changes to deliver scalable solutions for clients.
- Proficiently deployed AI/ML solutions on the Azure platform.

## EDUCATION

**BITS Pilani** (01/09/2019 - 01/04/2022)  
M.Tech, Data Science & Engineering

**RCC Institute of Information Technology (MAKAUT)** (01/08/2013 - 31/07/2017)  
B.Tech, Computer Science

**National English School** (01/04/2011 - 31/03/2013)  
ISC, Science

**National English School** (01/04/2009 - 31/03/2011)  
ICSE, Science

## PROJECTS

### Healthcare Consumer Analytics, Genpact (June 2022 - October 2023)

Data Scientist (Team Lead)

- The solution aims to optimize promotional data based on account-level requirements.
- Achieved a significant cost saving of \$2.3 million through the project.
- Demonstrated a high level of accuracy, with the model achieving over 90% accuracy across all products.

### Demand Sensing, Capgemini (August 2021 - June 2022)

Data Scientist (Team Lead & SME)

- The solution's objective is to customize and scale up the existing demand sensing tool according to the client's needs.
- Currently utilizing Pyspark for model training, resulting in a significant reduction of training time from 3 days to approximately 2 hours.
- Implementing re-design and optimization of machine learning and deep learning models to accelerate convergence and achieve a 5% improvement over the base model.
- Introduced a new best model scoring algorithm that enhances the best model selection process by 10%.
- Actively involved in client-level discussions and successfully managed a team of 15 members, including building impact analysis and risk assessment reports throughout the project duration.
- Delivered an enhanced Demand Forecast solution with a remarkable reduction of 60% in manual intervention. Worked with technologies such as Python, multiprocessing, multi-threading, and Pyspark.

### Demand Sensing, Capgemini (March 2020 - August 2021)

Data Scientist (Co-Team Lead)

- Developed an end-to-end Demand Sensing solution with a specific objective in mind.
- Created Python modules for data ingestion, processing, and forecasting pipelines.
- Developed Flask APIs for the user interface (UI), designed UI wireframes, and provided guidance to a team of 2 for building the frontend application.
- Utilized various models such as ARIMA, Exponential Smoothing, LSTM, MLP, XGBoost, Croston, Tbats, Random Forest, Moving Average, and Naive forecaster.
- Implemented hyper-parameter optimization using hyperopt, enabling potential hyper-tuning of models to improve accuracy.
- Utilized Dask to reduce computation time when necessary.

### Predictive Maintenance, Capgemini (December 2019 - March 2020)

Data Scientist

- The objective of the project was to predict the time of failure for a specific component of an electric vehicle.
- Conducted exploratory data analysis (EDA) on time-varying data, which included various features related to vehicle components. Identified the most important features and built a structured dataset to predict the time of failure within a specified time window provided by the client.
- Utilized XGBoost and Decision Tree models, and employed hyperopt for parameter tuning.
- Proposed an advanced methodology called Time Varying Cox Proportional Hazard, which predicts the probability of failure at a specific point in time.
- Created interactive visualizations using plotly and ipywidgets.

### Service Now Analytics, Capgemini (October 2019 - December 2019)

Data Scientist

- The objective of the project was to predict the volume of tickets to be generated in the near future and develop a customized routing algorithm for the client.
- Implemented LSTM and MLP models for ticket volume prediction.
- Developed a custom routing algorithm that achieved an accuracy of 80% on the validation dataset.
- Provided guidance to a team of 2 Power BI resources to create informative and visually appealing visuals for the client.

## PROJECTS

### Supply Chain Management, Capgemini (April 2019 - October 2019)

Junior Data Scientist

- The objective of the project was to develop an end-to-end supply chain pipeline that includes demand forecasting, data cleaning, and data wrangling for time series data.
- Implemented LSTM, MLP, Croston, ARIMA, and XGBoost models.
- Built both uni-variate and multi-variate models, incorporating social media data (open source) as the multi-variate feature. Extracted social media data and processed it to determine the sentiment of responses (Positive, Negative, Neutral) using Vader Sentiment Analysis.
- After model training, utilized Grid Search method to optimize parameters within a small parameter space and identify the best parameters for improved performance.

### Video Analytics Solution, Capgemini (October 2018 - March 2019)

Junior Data Scientist

- The objective of this internal proof of concept (POC) was to extract people's faces and tag the brands they are wearing.
- Implemented a process to extract and analyze images of individuals, identifying the brands displayed on their clothing.
- Utilized the YoLo v2 model to extract images of individuals, and employed the Inception v3 model to identify the specific brands being worn

## Trainee

- Company : Capgemini (April 2018 - July 2018)
- Company : TECTES (February 2018 - March 2018)
- Company : Ideation Technology Solutions Pvt. Ltd. (September 2017 - November 2017)
- Project : WiFi based PBX System, IBM Remote Mentoring Project (October 2016 - February 2017)
- Project : Tender Management Project, Globsyn Skills (June 2016 - July 2016)
- Project : Integration of MSSQL Server with Solr, IBM Remote Mentoring Project (February 2016 - July 2016)

## SKILLS

### PROGRAMMING

Python, Java Script.

### FRAMEWORKS & LIBRARIES

- Data Science : NumPy, Scikit-Learn, Pandas, Tensorflow, Keras, Scipy, OpenCV, statsmodels, nltk, hyperopt, MLOps, celery, faust, kafka, pyspark, GenAI (LLM).
- Visualization : Seaborn, Matplotlib, Plotly.
- Web : FastAPI, Flask.
- UI : DASH, streamlit.
- Database : MSSQL, MySQL, Redis, Postgres.

### SOFTWARES KNOWN

Anaconda, Spyder, PyCharm, VS Code, Jupyter Notebook, ML Azure, AWS Sagemaker, Google Colab, Power BI, Docker, Data Robot, AWS Lambda Function, AWS EBS, AWS Transcribe, AWS Comprehend.

## AWARDS

### Star Performer Award

Genpact

Performed multiple POC's in which 2 were successfully deployed at client's end, along with ontime deliveries with no escalation

(01/05/2023)

### GDSC 2020 (Global Data Science Challenge)

Capgemini

Attained in top 10% out of 1693 competitors globally.

### Super Team Award

Capgemini

Brought 6+ projects within the organization.

### GDSC 2021 (Global Data Science Challenge)

Capgemini

Attained 1st at global level.

### Innovation Award

Capgemini

Build an Innovative Solution for the organization.

### Rising Star Award

Capgemini

Capgemini 5 times ( in April 2019, September 2019, December 2020, January 2021 and December 2022) received for Customer delivery excellence.

## CERTIFICATIONS

### Generative AI with Large Language Models

DeepLearning.AI Issued November, 2023

### ML Practitioner Certificate

Dataiku Academy Issued July, 2023

<https://verify.skilljar.com/c/2hph8ogspvy5>

### Core Designer Certificate

Dataiku Academy Issued July, 2023

<https://verify.skilljar.com/c/4ucupx9wj33r>

### Neural Networks and Deep Learning

DEEPLARNING.AI Issued on January, 2022

### Agile Software Development

COURSERA Issued on August, 2021

### Convolutional Neural Networks in TensorFlow

DEEPLARNING.AI Issued on September, 2021

### Natural Language Processing in TensorFlow

DEEPLARNING.AI Issued on June, 2020

### MLOps Practitioner Certificate

Dataiku Academy Issued August, 2023

<http://verify.skilljar.com/c/7djsjp74vdvj>

### Advanced Designer Certificate

Dataiku Academy Issued July, 2023

<http://verify.skilljar.com/c/weie95jezjyz>

### Build Basic Generative Adversarial Networks (GANs)

DeepLearning.AI Issued June, 2022

### Sequences, Time Series and Prediction

DEEPLARNING.AI Issued on September, 2021

### Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning

DEEPLARNING.AI Issued on September, 2021

### Machine Learning with Python

IBM Issued on June, 2021

### Python for Data Science

IBM Issued on May, 2019

## LANGUAGES

### English

8



### Hindi

8

