

# SUMIT KUMAR MISHRA

## Data Scientist

[Github](#)

[LinkedIn](#)

### EDUCATION

Bachelor of Technology. Computer Science and Engineering

Bhawani Shankar Anangpuria  
Institute of Technology and  
Management

2015 – 2019  
Faridabad, Haryana

### Skills

#### LANGUAGES

• Python • SQL (Database Language)

#### TECHNICAL TOOLS

• Statistical Data Analysis • Logistic Regression • Random Forest • Decision Tree  
• Machine Learning Algorithms • NLP (Natural Language Processing) • Mathematics • Data visualization • Predictive Modeling • Data Structures and Algorithms (DSA) • Deep Learning • Forecasting

#### LIBRARIES

• TensorFlow • SciKit-Learn • Pandas • Numpy • matplotlib • seaborn

#### Dashboards

• Data Studio • Tableau • PowerBI

### Soft Skills

• Analytical • Problem-Solving Capability • Written and Oral Communication • Time Management • Proactive • Confident

## Aspiring Data Scientist

- 2 years experienced Software Engineer looking for transition to Data Scientist Role. With interest in analyzing raw data and statistics, brings excellent communication abilities with good knowledge of building machine learning models with deployment and creating dashboards with a story to tell.
- Seeking a role in this company where I can pose and answer questions with quantitative-driven insights.

### WORK EXPERIENCE

#### Software Engineer

**GlobalLogic India** 2021-current

Gurugram Haryana

- Designing and Implementing Complex ETL pipeline to ingest data and serve data in Google Search.
- Employed data cleansing methods, significantly Enhanced data quality.
- Generated detailed studies on potential third-party data handling solutions, verifying compliance with internal needs and stakeholder requirements.
- Conducted daily interaction with the Google Clients to understand project requirement.
- Explained data results and discussed how best to use data to support project objectives.
- Created highly scalable dashboard and workflows that help monitor the pipeline thereby reducing the failure of the pipeline. Tuned systems to boost performance.
- Utilized advanced querying, visualization and analytics tools to analyze and process complex data sets.

### PROJECTS

#### • Movie Recommender System

**Primary goal:** Create a recommender engine that predict list of relevant movie based on past data.

**Solution:** Created a content based movie recommender system, based on the similarity of movie attributes using cosine similarity.

**Project Link**

#### • Book Recommender System

**Primary goal:** Create a recommender engine that predict list of relevant books based on past data.

**Solution:** Created a collaborative based book recommender system, based on the similarity of movie attributes using cosine similarity.

**Project Link**

- Fake News Detection System

**Primary goal:** Devise a system that detects fake news from a large set of new articles.

**Solution:** Used Logistic Regression ML model to train the data that predicts whether the news is fake or not.

**Project Link**

- ED Analysis of Fortune 500 Companies

**Primary goal:** Perform data analysis of the top unicorn companies

**Solution:** Analyzed the data using visual techniques. Discovered trends, patterns, and checked assumptions with the help of statistical summary and graphical representations.

**Project Link**