

Koushik Khan

S&P Global Inc., Bangalore, India

Phone: (+91) 7501-923-251 | Email: koushikkhan38@gmail.com | Portfolio: <https://koushikkhan.github.io/>

SKILLS

PROGRAMMING LANGUAGES

- R
- Python
- Shell Scripting
- Julia (a novice)

ALGO. USED IN PROJECTS

- Regression & Classification
- Tree Based Modeling
- Recommender system
- Text Mining and NLP
 - Text cleaning pipeline
 - Text classification
 - Sentiment detection
 - Text search
 - Topic detection/clustering

FRAMEWORKS & TOOLS

- **Linux** (Debian|Ubuntu|Fedora)
- **MLOps** (Docker|GitHub Actions)
- **Version Control** (Git)
- **Dashboard** (R Shiny)
- **Distributed Systems** (PySpark)
- **Databases** (MySQL|MongoDB)
- **AWS Services** (S3|EC2)
- **Modelling Frameworks** (Need basis)
 - Web scrapping: BeautifulSoup
 - Data cleaning: Numpy, Pandas
 - ML models: SkLearn
 - DL models: PyTorch & Tensorflow

PERSONAL PROJECTS

- Described on my [portfolio](#)

EDUCATION

M.SC, STATISTICS

VISVA BHARATI UNIVERSITY

June 2015 | West Bengal, India

Thesis: "Joint Modeling of Longitudinal and Time to Event Data"

CGPA: 8.43

B.SC, STATISTICS

VISVA BHARATI UNIVERSITY

June 2013 | West Bengal, India

Graduated with Honors

CGPA: 8.26

CERTIFICATIONS

- Microsoft Certified: Azure Data Scientist Associate
- Microsoft Certified: Azure AI Fundamentals
- Microsoft Certified: Azure Fundamentals

EXPERIENCE

S&P GLOBAL | SR. DATA SCIENTIST

Aug 2021 – Present | Bangalore, India

- **Working With Shipping Data As Part of Product Management**
 - **Product management:** regular maintenance, bug fixes & addition of new features to cater the needs of stakeholders and clients
 - **Team management:** leading a separate team for external data acquisition, developing ETL pipelines and automation tools to onboard external data
 - **Data analysis:** analyzing mainstream shipping data for vessel tracking, cargo grade analysis, cargo volume analysis etc.
 - **Modelling:** developing models to better understand the pattern in the shipping data, predict certain business metrics and answer client queries to enrich the overall quality of the product

IBM | DATA SCIENTIST

Oct 2016 – July 2021 | Bangalore, India

- **Automated Ticket Classification System using ML**
 - worked on developing an automated system to analyze and classify service tickets that used to parse a large amount of textual data coming from XML and CSV files
 - worked single-handedly on packaging the application and deploying the same on a kubernetes cluster
- **Text Parsing Pipeline for IBM Watson API**
 - contributed in developing text cleaning pipeline for IBM Watson sentiment detection API
- **Serving Search Queries Using a Small Scale Search Engine**
 - worked on developing a small scale information retrieval system for one of our banking clients for the FAQ section on their main website
- **Marketing Mix Modelling to Analyze Sales and Advertising Effects**

BCAUSE ENTERPRISE PVT. LTD. | JUNIOR DATA ANALYST

Oct 2015 – Aug 2016 | Noida

- **Developing a simulator to calculate likelihood of packet flow through different channels in an electronic device**
- **Working with the database team for database maintenance and adhoc analyses**
- **Predicting the names of human body parts from a corpus of digitized prescriptions using text classification**