

SAI SUHAS VARIK

Sheffield, United Kingdom

saivarik15@gmail.com | +447867091657

LinkedIn: <https://www.linkedin.com/in/sai-varik>

PROFILE SUMMARY:

Experienced and ambitious Data Science professional with 4.5 years of hands-on experience in statistical analysis, data modeling, and programming. Proficient in data handling, cleaning, preprocessing, visualization, and machine learning models. Skilled in Python, R, SQL, AWS, SPSS, KNIME, Databricks, RStudio, Tableau, MATLAB and Oracle Apex. Passionate about leveraging data to solve complex problems and continuously seeking opportunities to enhance skills and knowledge. Demonstrated leadership and relationship-building abilities, fostering collaborative environments to drive innovative solutions.

EDUCATION:

The University Of Sheffield, MSc in Data Science, United Kingdom - Full-time, Sept 2022 – Sept 2023

Coursework: Data Analysis, Introduction to Data Science, Data Visualization, Big Data Analytics, Business Intelligence, Data Mining, Database Design, Data and Society, Amazon Web Services (AWS), Research Methods and Dissertation Preparation.

Industry Project: Sentiment Analysis and Topic Modeling of South Yorkshire Mayoral Combined Authority (SYMCA) Tweets to Measure Success and Improve Data-Driven Feedback.

Grade: Distinction (predicted)

Visvesvaraya Technological University, B.E. Electronic and Communication, India – Full-time, Aug 2015 – Jun 2019

Coursework: Computer Communication Networks, Theory and Coding, Programming in C & C++, Microprocessor, Network Analysis, Digital Image Processing, Management & Entrepreneurship Development, MATLAB.

Major Projects: Waste Segregation system using Smart Dustbin, Wireless Patient Health Monitoring System.

Grade: Merit

WORK EXPERIENCE:

University of Sheffield, Graduate Teaching Assistant, Information School, United Kingdom, Sept 2023 – Present

Facilitated student learning across diverse subjects, bridging theory and practice, utilizing industry-standard tools and fostering active engagement. These roles collectively enriched students' understanding of data analysis, information modeling, and digital business, equipping them for the dynamic demands of their respective fields.

Data Analysis (G6 teaching support):

- Provided students with a comprehensive understanding of statistical data analysis, covering a wide range of techniques from descriptive statistics to advanced multivariate methods.
- Utilized industry-standard tools such as IBM SPSS and real-world examples to bridge the gap between theoretical knowledge and practical application in data analysis.

Information Systems Modelling (G6 & G7 teaching support):

- Introduced students to a diverse array of modeling techniques, including soft systems analysis, structured systems analysis methodologies, and object-oriented approaches like RUP/UML, equipping them with a versatile skill set.
- Conducted engaging seminars, tutorials, and small group activities to actively involve students in the learning process, ensuring they comprehensively grasped complex modeling methodologies.

Digital Media and Society (G6 marking support):

- Offered a holistic education experience by covering both theoretical and practical aspects of digital business, enabling students to understand, adapt to, and succeed in the dynamic digital business landscape.
- Explored the latest trends and business models adopted by e-commerce companies, ensuring students remained up-to-date with contemporary industry practices and equipped them with theoretical practical skills in the digital economy.

South Yorkshire Mayoral Combined Authority (SYMCA), Social Media Consultant, United Kingdom, Jan 2023–Sept 2023

- Led a collaborative project with South Yorkshire Mayoral Combined Authority to optimize social media presence.
- Developed a comprehensive campaign strategy with the Data Services and Research Manager.
- Utilized Python for web scraping and data gathering, ensuring accuracy through robust data cleaning and preprocessing.
- Managed high – volume social media data and also demonstrated expertise in data storage and retrieval of datasets.
- Applied Exploratory Data Analysis (EDA) and advanced Natural Language Processing (NLP) techniques for sentiment analysis, and topic modelling including Latent Dirichlet Allocation (LDA) and language model based on BERT.
- Delivered actionable insights and comprehensive reports, leading to heightened online visibility for SYMCA.

Infosys Limited, Senior System Engineer, India – Full-time, Jan 2020 – Aug 2022

- Managed end-to-end product development of Citibank's Instant Payment (IP) Processing System in Australia, leading cross-functional teams and achieving outstanding results, resulting in being awarded the Star Performer Award.
- Developed and implemented CI/CD pipelines using Bitbucket, Teamcity, and Udeploy, improving product efficiency.
- Executed Azure Lift and Shift migration by enhancing report performance through SQL stored procedures and Azure Data Factory (ADF) pipelines, as well as automating ETL processes using SSIS, Scala Jobs, and Python code adjustments for seamless execution in Azure Data Bricks.
- Developed a training project that involved creating a dashboard for construction equipment using Power BI, Angular JS
- Developed framework to seamlessly ingest data from diverse file sources (CSV, JSON, XLSX, DBMS) with varying schemas, efficiently transforming and storing it in MongoDB.
- Profusely worked on Version control tools like GIT and Jira Software for project Management and issue tracking.
- Experienced in working on Agile operations process and tools area, and arranged Confluence calls every week.

- Testing the code using Behaviour Driven Development (BDD) and ensuring proper documentation of results.
- Worked with a team of diversified web designers, back-end developers and UI designers to meet project goals.
- Trained and certified in Blockchain to create smart contracts using Ethereum platform.

Railwire Corporation of India Ltd, Supervisor, India – Part-time, Jun 2019 – Dec 2019

- Managed customer data entry, plan allocation, and databases for multiple OLT devices, also ensured efficient bandwidth allocation based on client utilization, provided technical leadership to the team members to track defects.

Qspiders CampusConnect, Intern, India – Internship, Jan 2019 – Feb 2019

- Completed an internship involving Python programming for various problem statements and real-world scenarios, effectively developing solutions and addressing day-to-day challenges.

Goa Shipyard Limited, Intern, India - Internship, July 2018 – Aug 2018

- Demonstrated proficiency in data entry and inventory management and ensuring accurate tracking of materials.

Synegra Ems Limited, Intern India – Internship, Jan 2018 – Feb 2018

- Managed stock inventory and performed data entry for a variety of communication devices including routers, modems, and other equipment, ensuring accurate record-keeping and efficient stock control.

CERTIFICATIONS:

- Microsoft Certified **Azure Data Engineer Associate** and **Azure AI Fundamentals** in 2023
- Certified by Infosys as **Big Data Developer, Agile Developer** in 2022 and **Python Developer, SQL Developer** in 2020
- Oracle Certified **Java Professional and Associate** certified by Oracle University SE 8 Programmer in 2021

PROJECTS:

Research Project: Analysis of the Impact of Hormonal Changes on Rheumatoid Arthritis

Tools: IBM SPSS Statistics

- Utilised the Wave 7 English Longitudinal Study of Ageing (ELSA) Data Set.
- Performed data extraction, cleaning, preprocessing, management, and visualization using IBM SPSS Statistics 28.0.0.0.
- Employed descriptive statistics to summarize the data and assess central tendency and variability.
- Applied multiple variable regression, Kolmogorov-Smirnov test, Chi-squared test and identified correlations.

Research Project: Multivariate Regression Analysis of Overseas Passenger Responses to Covid-19 Parameters

Tools: R Studio

- Conducted multivariate regression analysis on overseas passenger responses to COVID-19 parameters.
- Utilised data from multiple sources (Our World in Data COVID-19 and Office for National Statistics dataset) using R.
- Cleaned and preprocessed the data including handling missing values, Converting data types, normalizing the data, aggregating and joining the datasets to examine the significance of COVID-19 testing and perception of safety.
- Explored correlations and effect sizes using Pearson coefficient and multivariate regression techniques.

Research Project: Exploring Job-Technology Skill Relationships

Tools: Databricks, DataFrames, RDD, SQL, Python, Map Reduce, HDFS, Hadoop, Apache Spark

- Integrated data from diverse sources (csv and O*NET) to examine the connection between job descriptions and sought-after technological skills and executed data manipulation techniques to harmonize the datasets.
- Utilised DataFrames, RDD, SQL, and MapReduce algorithms in Databricks for efficient data analysis, processing, and complex transformations, resulting in improved processing times and enhanced performance.
- Conducted skill analysis to detect patterns and correlations among job skills.

Research Project: Binary Classification on Titanic Dataset

Tools: Knime, Machine Learning Models

- Researched and implemented automatic feature selection algorithms and binary classifiers (Naive Bayes, Random Forest, Decision Tree) in Knime, also evaluated performance using accuracy, AUC, Matthew's Correlation Coefficient (MCC).
- Identified important socio-economic and socio-demographic factors that influenced survival.
- Offered valuable observations on the significance of features and their influence on predictions regarding survival.
- Improved comprehension of the Titanic dataset by conducting binary classification analysis.

Industry Project: Vodafone Plc Company Report

Tools: PowerBI, Mintel, FAME database, The Financial times, financial reports analysis, social media analysis

- Evaluated financial data from FAME, The Financial times, Mintel and the annual report from company's website.
- Implemented SWOT and PESTEL analysis to understand business environment.

SKILLS:

Professional: Analytical thinker, Good communication skills, Time management, Leadership qualities, Adaptive, Meticulous

Coding Languages: Python, R, SQL, Java, Scala, HTML, CSS, C, C++

Algorithm: Classification algorithm, Regression algorithm, Deep-learning, Convolutional Neural Network (CNN)

Databases: MySQL, Oracle Database, PostgreSQL, Azure Synapse Analytics, MongoDB, Mintel, The Financial Times, FAME

Data Analysis: T-SQL, JQuery, SSIS, PowerBI, Kafka

Version Control: GIT, Jira, CI/CD pipelines | **BDD Tools:** Cucumber | **Cloud:** Azure Databricks, Azure Data Factory, AWS

Platform: RStudio, SPSS, KNIME, Databricks, Oracle Apex, Eclipse, Visual Studio, MATLAB, Canva, Tableau, MS Office