Sarvesh Suresh

Data Science Professional
Post Graduate in Statistics

Email : sarveshstats@gmail.com
Phone : +91-9080009221

Academic Qualifications

Year	Degree/Certificate	Institute	CPI/%
2021-2023	M.Sc Statistics	Indian Institute of Technology, Kanpur	7.6/10
2018-2021	B.Sc Statistics	Loyola College, Chennai.	96.8%

Work Experience

- **Data Science Analyst, Accenture Strategy and Consulting -Applied Intelligence.** (*Jul'23 onwards*)
  - Prepared a **financial database** for a potential Banking and Finance Company for it to be used as input for the predictive part of an internal widget.
  - Built a feedback loop using Python for a **GenAI-based General Ledger records tagging** project.
  - Set up the structured (snowflake) and unstructured (pdf files) database for an internal GenAI accelerator using python based preprocessing.
  - **Prompt Engineering** for a GenAI based-Co-Pilot using the GPT-4 API in Python, which included- Summary creation, Identifying Key ideas from text, creating graphs/pictures, fitting text into structured data.
  - **Data Analytics for Product Segmentation** for a Fortune 500 global branded food company **(since December'23)**
    - Extensive pre-processing using **GCP-BQ stored procedures and Dataiku flows**
    - ML-based segmentation combined with heuristic decision rules.

Internships

- **Data Science Analyst Intern, Accenture-Applied Intelligence.** (*May'22- Jul'22*)
  - Built and published a Python Package called **fairness-audit**, that contains modules that upon receiving the raw data set, target variable and sensitive variable, automate the pre-processing, model-building and fairness evaluation tasks of an ML Pipeline.
  - Statistical testing of proportions across various metrics of model performance are carried out between the different levels of the sensitive class in order to evaluate fairness.
- **Data Science Intern, Chariot Beach Resorts, Mahabalipuram.** (*Mar'21- Apr'21*)
  - Carried out parametric tests on ratings-data to compare the performance of the resort in terms of **customer satisfaction** computed year-wise, season-wise and factor-wise.

Achievements and Qualifications

- Passed the **CS1 (Core Statistics-1)** paper conducted by the **Institute of Actuaries, India**.
- Received the **Gold Medal** and **Scholarship**, for exceptional academic performance during the period 2018-2021.

Key Projects

- **Product Segmentation (Client Project, Accenture)** (*December'23-present*)
  - **End-to-End Ownership**: Worked directly under a Senior Manager, with no other technical team members, handling all aspects of the product independently from data engineering to deployment.
  - **Enhanced Data Join Accuracy**: Improved raw table join rate from 45% to 99.7% by analyzing data structures, refining rollup levels, and identifying correct join keys.
  - **Built Scalable Pipelines**: Developed Python/Dataiku pipelines to clean, roll, and join multiple datasets from GCP, ensuring seamless data integration.
  - **Optimized Data Preprocessing**: Converted preprocessing steps before clustering into stored procedures in BigQuery, streamlining execution and improving efficiency.
  - **Implemented Segmentation Workflows**: Set up clustering and thresholding pipelines in Dataiku, ensuring segmentation followed industry best practices.
  - **Automated Service Level Targeting**: Designed and implemented automated service level target assignment for wholesale segmentation, reducing manual intervention.
  - **Resolved Pipeline Breakage Issues**: Modified the thresholding process to handle edge cases, ensuring a robust and fail-safe segmentation pipeline.

  - **Led Cross-Team Coordination**: Collaborated with Accenture Data Engineering Client Data Science teams to promote pipelines from Dev → TST → PRD using Liquibase.
  - **Adapted to Changing Requirements**: Quickly modified pipelines to accommodate last-minute changes (switching to Product-DC level output) under tight deadlines.
  - **Unified Retail Wholesale Pipelines**: Created a common module to standardize final segment assignments across both business units.
  - **Developed a Scalable Framework**: Extended the approach to Wholesale, designing a parallel yet distinct pipeline tailored to its unique processing needs.
- **Clustering on NIFTY (For course: MTH552A)** (*April'22*)
  - This project was undertaken to show that the NIFTY 50 index can be trimmed using K-Means Clustering based on Financial Variables to produce superior returns (alpha) over the original index.
  - We use survivor-bias-free clustering-based back-testing to estimate the potential future alpha that could be generated.
- **Project entitled audit of Endoscopic Mucosal Resection of duodenal adenomas: Factors influencing success and reducing complications.** (*Dec'21-Jun'22*)
  Guide: Dr Jayapal Ramesh, Advanced Endoscopist, Royal Liverpool and Broadgreen University Hospitals NHS Trust, Honorary Senior Lecturer, University of Liverpool.
  - Providing statistical help for the project by carrying out statistical tests of significance in order to confirm the dependence of the probability of a successful procedure on the various physiological and procedure-specific variables
  - Help in writing the Statistics part of the manuscript.
- **Historical Fundamental Screener (Self-Project)** (*May'21-Present*)
  - The HFS is a stock research tool that lets the user filter stocks based on financial ratios, rank them by a certain ratio, select the size of your portfolio and see how the strategy would have performed over the last 10 years. The financial data used for this app was scraped using BeautifulSoup
  - Currently working on estimating and mitigating the inherent survivor bias.
- **Sports-Network-Optimization using Survey Analysis (For Course : 18UST6MS01)** (*Apr'21*).
  - The objective of this survey and its analysis was to put myself in the shoes of a **Sports Broadcasting Network** to decide on the events to cover over a period of one year, subject to the budget as well as tournament-related constraints in order to maximize the Revenue
  - Calculated probability-based custom popularity metrics and used **Linear Programming** to obtain the optimized results
- **Analysis of the socio-economic factors affecting the spread of COVID 19 in India.** (*May'20*)
  Guide: Dr Martin.L.William, Former Head of the Department of Statistics, Loyola College, Chennai.
  - Implemented **Linear Regression** techniques like Ordinary/Weighted Least Squares, Lasso and Ridge, and analyzed the **statistical significance** of the various state-wise factors on the proportion of positive cases observed.
  - Included the interaction effects for 'time' and other relevant variables and came up with **day-wise projections** for the number of new COVID cases in Tamil Nadu.

Technical Skills

- **Languages:** R, Python, SQL
- **Software:** SPSS, Microsoft-Excel, Dataiku, GCP
- **Skills:** Data Analysis, Stock Market- Backtesting, Web Scraping.