

# Sanjith Ganesh

Phone: +1 (908)672-2790 | Email: [gsj2442@gmail.com](mailto:gsj2442@gmail.com) | LinkedIn: [sanjith-ganesh](#) | Git: [Sanjith](#)

## Education

<b>Rutgers University, New Jersey</b> <i>M.S in Data Science, Major - Statistics</i>	Expected - May 2026
<b>Relevant Coursework:</b> Probability and Statistical Inference, Regression Analysis, Machine Learning, Data Structure, Data Wrangling, Statistical Software, Data Mining	
<b>SSN College of Engineering</b> , Chennai, India <i>B.E in Computer Science and Engineering</i>	June 2024
<b>Relevant Coursework:</b> Python, Computer Networks, DBMS, Computer Architecture, Operating Systems, Compiler Design, C, Java Programming, Web Designing, Intro to AI and ML	

## Experience

<b>CSC Global, Wilmington, Delaware</b> <i>Data Science and Business Intelligence Intern</i>	May 2025 - Aug 2025
<ul style="list-style-type: none"><li>Built an automated PDF validation tool using <b>Python (OpenCV, SSIM, PyMuPDF, and Pillow)</b> to detect pixel-level mismatches and auto-capture error screenshots locally.</li><li>Developed a fuzzy-matching pipeline with <b>Oracle SQL, Python (RapidFuzz), and Alteryx</b>.</li><li>Created an Advance Payment workflow integrating <b>Power BI, Power Apps, and Power Automate</b>, automating email generation and SharePoint Excel synchronization.</li><li>Automated <b>SAP Collections</b> by exporting Power BI visuals to <b>Excel via JSON array logic</b>, reducing manual multi-row updates.</li></ul>	
<b>SAAFE Dashboard Account Aggregation Pvt. Ltd., Chennai, India</b> <i>Data Analyst Intern</i>	May 2023 - Sep 2023
<ul style="list-style-type: none"><li>Developed Account Aggregator modules using <b>Java, Python (Pandas, NumPy)</b>, and <b>MongoDB</b> to analyze banking datasets and extract customer insights.</li><li>Built interactive dashboards in <b>Power BI and Tableau</b> for visualizing customer adoption trends and Technological Service Provider (TSP) activity.</li><li>Designed and deployed a responsive Frontend Website using <b>HTML, CSS, and JavaScript</b>, ensuring clean UI/UX and real-time data integration from backend APIs.</li><li>Engineered System Flow Diagrams using <b>Lucidchart, Draw.io, and StarUML</b> to model data pipelines</li></ul>	
<b>Nutriton Organics, Ahmedabad, India</b> <i>Business and R&amp;D Data Analyst Intern</i>	Jan 2023 - May 2023
<ul style="list-style-type: none"><li>Analyzed <b>chemical and production data</b> using <b>Python (Pandas, NumPy)</b> to derive insights for R&amp;D and business decisions.</li><li>Improved <b>billing and sales workflows</b> through process analysis and efficiency tracking.</li><li>Applied <b>SQL, Power BI, Excel, and Jupyter</b> for data visualization and reporting.</li></ul>	

## Technical Domain and Other Skills

**Core Domain Expertise:** Data Analysis, Business Analysis, Business Management, Risk Prediction & Management, HR Management, AR/VR Basics, Process & Supply Chain Management, NLP, Causal Inference, Data Engineering (ETL & pipelines) and Visualization

**Programming Languages:** Python, Java, C, C++, SQL, R, DAX, M (Power Query basics), Typescript

**Web Design & Development:** HTML, CSS, JavaScript, Bootstrap, Flask, NodeJS, NestJS, React, Flutter

**Data Science & ML Tools:** Alteryx, scikit-learn, TensorFlow (basics), NumPy, Pandas, , PyTorch, LoRA/PEFT, Fine Tuning NLP

**Visualization & Analytical Tools:** Power BI, Tableau, Microsoft Excel, Jupyter Notebook, Google Analytics

**Databases:** MySQL, MongoDB, Oracle, PostgreSQL, Firebase

## Research Project

---

<b>CLEF Recognition using Image Processing</b>	May 2022
<ul style="list-style-type: none"><li>Built and labeled a <b>custom flower image dataset</b> for training and validation.</li><li>Implemented an <b>image recognition model</b> using <b>Python (OpenCV, TensorFlow, scikit-learn)</b> for flower type classification.</li><li>Enhanced recognition accuracy through <b>color space transformations (RGB-HSV), edge detection, image segmentation (K-means, thresholding)</b>, and feature extraction (<b>HOG, CNN layers</b>).</li></ul>	
<b>Image Processing to Detect Natural Calamity Victims from Social Media</b>	March 2023
<ul style="list-style-type: none"><li>Collected and preprocessed <b>social media images</b> using <b>Instagram and Facebook APIs</b>, applying <b>keyword filtering and temporal sampling</b> for dataset creation.</li><li>Trained a <b>CNN model</b> in <b>TensorFlow and Keras</b> with <b>transfer learning and data augmentation</b> to classify disaster-related visuals.</li><li>Applied <b>image segmentation and edge detection</b> using <b>OpenCV</b> to enhance feature localization and reduce false positives</li><li>Built a <b>Flask-based web module</b> integrating <b>real-time detection and mapping</b>, placing <b>5<sup>th</sup></b> at the <b>All India National Hackathon</b>.</li></ul>	
<b>Automated Code Integrity Checker (Plagiarism + Grading Evaluation Tool)</b>	March 2024
<ul style="list-style-type: none"><li>Developed an LMS-integrated tool using <b>Python, Flask, and MySQL</b> to automate <b>plagiarism detection, syntax validation, and grading</b> for student assignments.</li><li>Compiled a 15-year dataset of submissions and applied <b>AST-based code comparison, tokenization, and Levenshtein distance metrics</b> for similarity analysis.</li><li>Implemented <b>Rule-based grading algorithms</b> with <b>regex parsing and error pattern detection</b> for code evaluation.</li><li>Approved by the <b>Computer Science Board of SSN College of Engineering</b> as the official grading tool; research paper under review at <b>IJSRET</b>.</li></ul>	

## Academic & Technical Projects

---

- Database System for Medical Health Center (May 2022)**: Built a medical database using **MongoDB** and **Java Framework** for intercollege data management.
- IDS and IPS Detection using Networking (Jul 2023)**: Implemented a **network security model** to detect malicious events using IDS/IPS protocols.
- Social Network Visualization using GEPHI (Nov 2024)**: Analyzed collaboration networks via **Gephi**, applying metrics like centrality and community detection.
- Growth Mindset Intervention Causal Impact Analysis (ATE Estimation) (Aug 2024)**: Estimated the causal effect of a “growth mindset” nudge on student achievement by modeling treatment assignment
- Subsampling vs Ridge Regularization (Theoretical + Empirical Study) (Mar 2025)**: Studied how **subsample** based linear ensembles compare to **ridge (L2) regularization** in high dimensional settings using controlled simulations in R.
- News Headline Classification (AG News) (Sep 2025)**: Benchmarked TF-IDF + **Logistic Regression/SVM** baselines against **DistilBERT** and **RoBERTa**, and evaluated full fine-tuning vs parameter efficient **LoRA + dropout regularization** with comparative metrics and error analysis across all models.
- Exponential Family Embeddings Comparison (Dec 2025)**: Compared **P-EMB, Additive/Downweighted P-EMB, and PCA/HPF baselines** to evaluate embedding quality and model fit on our dataset.
- Attendance Management System (Advanced DBMS) (Dec 2025)**: Built a **role-based web app** (Student/Teacher/Admin) with a **PostgreSQL** backend to manage enrollments and attendance workflows, including secure login, CSV import/export, and **dynamic dashboards**.
- Movie Data Warehouse Pipeline (Web Scraping + ETL) (Dec 2025)**: Built an end-to-end pipeline to scrape, clean, normalize, and load movie metadata into a fact-dimension warehouse for fast SQL analytics.



