# Sanjith Ganesh

Phone: +1 (908)672-2790 | Email: gsj2442@gmail.com | LinkedIn: sanjith-ganesh
Git: Sanjith | Portfolio: SanjithGanesh

## Education

**Rutgers University,** *New Jersey* Expected - May 2026
*M.S in Data Science,* ***Major -*** *Statistics*
**Relevant Coursework:** Probability and Statistical Inference, Regression Analysis, Machine Learning, Data Structure, Data Wrangling, Statistical Software, Data Mining

**SSN College of Engineering,** Chennai, India June 2024
*B.E in Computer Science and Engineering*
**Relevant Coursework:** Python, Computer Networks, DBMS, Computer Architecture, Operating Systems, Compiler Design, C, Java Programming, Web Designing, Intro to AI and ML

## Experience

**CSC Global,** *Wilmington, Delaware* May 2025 - Aug 2025
*Data Science and Business Intelligence Intern*

- Built an automated PDF validation tool using **Python (OpenCV, SSIM, PyMuPDF, and Pillow)** to detect pixel-level mismatches and auto-capture error screenshots locally.
- Developed a fuzzy-matching pipeline with **Oracle SQL, Python (RapidFuzz), and Alteryx**.
- Created an Advance Payment workflow integrating **Power BI, Power Apps, and Power Automate**, automating email generation and SharePoint Excel synchronization.
- Automated **SAP Collections** by exporting Power BI visuals to **Excel via JSON array logic**, reducing manual multi-row updates.

**SAAFE Dashboard Account Aggregation Pvt. Ltd.,** *Chennai, India* *May 2023 - Sep 2023*
*Data Analyst Intern*

- Developed Account Aggregator modules using **Java, Python (Pandas, NumPy)**, and **MongoDB** to analyze banking datasets and extract customer insights.
- Built interactive dashboards in **Power BI and Tableau** for visualizing customer adoption trends and Technological Service Provider (TSP) activity.
- Designed and deployed a responsive Frontend Website using **HTML, CSS, and JavaScript**, ensuring clean UI/UX and real-time data integration from backend APIs.
- Engineered System Flow Diagrams using **Lucidchart, Draw.io, and StarUML** to model data pipelines

**Nutriton Organics,** *Ahmedabad, India* *Jan 2023 - May 2023*
*Business and R&D Data Analyst Intern*

- Analyzed **chemical and production data** using **Python (Pandas, NumPy)** to derive insights for R&D and business decisions.
- Improved **billing and sales workflows** through process analysis and efficiency tracking.
- Applied **SQL, Power BI, Excel, and Jupyter** for data visualization and reporting.

## 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

**CLEF Recognition using Image Processing** (🔗) May 2022

- Built and labeled a **custom flower image dataset** for training and validation.
- Implemented an **image recognition model** using **Python (OpenCV, TensorFlow, scikit-learn)** for flower type classification.
- Enhanced recognition accuracy through **color space transformations (RGB–HSV), edge detection, image segmentation (K-means, thresholding), and feature extraction (HOG, CNN layers)**.

**Image Processing to Detect Natural Calamity Victims from Social Media** (🔗) March 2023

- Collected and preprocessed **social media images** using **Instagram and Facebook APIs**, applying **keyword filtering and temporal sampling** for dataset creation.
- Trained a **CNN model** in **TensorFlow and Keras** with **transfer learning and data augmentation** to classify disaster-related visuals.
- Applied **image segmentation and edge detection** using **OpenCV** to enhance feature localization and reduce false positives
- Built a **Flask-based web module** integrating **real-time detection and mapping**, placing **5th at the All India National Hackathon**.

**Automated Code Integrity Checker (Plagiarism + Grading Evaluation Tool)** (🔗) March 2024

- Developed an LMS-integrated tool using **Python, Flask, and MySQL** to automate **plagiarism detection, syntax validation, and grading** for student assignments.
- Compiled a 15-year dataset of submissions and applied **AST-based code comparison, tokenization, and Levenshtein distance metrics** for similarity analysis.
- Implemented **Rule-based grading algorithms** with **regex parsing and error pattern detection** for code evaluation.
- Approved by the **Computer Science Board of SSN College of Engineering** as the official grading tool; research paper under review at *IJSRET*.

## 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- Cloud Based (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 **subsampling** 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 Recommendation Data-Warehouse Pipeline (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.