

Sanchit Goel

goelsanchit.com • s5goel@ucsd.edu • +91 9717252475
github.com/sanchitgl • in -sanchitgoel

EDUCATION

University of California San Diego | CALIFORNIA, USA

Masters in Data Science | 2023-2025

Ashoka University | DELHI, INDIA

P.G.D in Advanced Studies and Research | 2019 - 2020

Minors in Computer Science | CGPA: 3.6/4 | **Dean's List**

Ashoka University | DELHI, INDIA

B.A. (Hons.) Economics | 2016 - 2019

Major in Economics | Major GPA: 3.4/4

RESEARCH EXPERIENCE

Research Assistant | DR. KANIKA MAHAJAN, ASHOKA UNIVERSITY

May 2020 – August 2020 | Delhi, India

- Worked as an RA on '**Words matter: Gender, Jobs and Applicant Behaviour [AAEA]**', to examine how implicit gender preference in Job ads leads to differential labor market outcomes. Used NLP techniques - TF-IDF, Topic modeling and Freq-dist to study the difference between the characteristics of jobs offered to males and females.
- Worked as an RA on '**Covid 19 and Supply chain disruption [IZA]**'. Computed the distances between India's production zones and retail centres using Google maps API to study the impact of distance in causing supply chain disruptions.

Research Assistant | CENTRE OF ECONOMICS & DATA ANALYSIS, ASHOKA UNIVERSITY

August 2021 – October 2021 | Delhi, India

- Worked with **Dr. Ashwini Deshpande** and **Dr. Joshua Merfeld** on the working paper '**Decentralization and Program Implementation**' to study the impact of district splits on the effectiveness of government program implementation.
- Prepared the dataset of over 800 districts and 7000 blocks and identified district splits in India from 2011-20 by computing block level changes year over year.

PROFESSIONAL EXPERIENCE

Data Scientist | FRACTAL ANALYTICS

Oct 2020 – Present | Gurugram, India

Awarded **STAR** and **SPOT Award** for commitment and leadership in FY 2021-22

- **Content Tagging Automation Platform**: Working with **Merck**- Building an intelligent NLP platform capable of identifying features such as purpose, key message and tone of marketing content. **Led the Data science effort for exploring Large language models and finalized Roberta, improving our solution's f1-score by over .12.**
- **Propay.bot**: E2E solution that processes and verifies health insurance claims (PA and FWA contracts) using NLP models - Spacy, PYSBD and Regex. **In process of adoption by top health insurance firms - United Health & Cigna.**
- **Invoice Information Extraction**: E2E solution that streamlines and automates Invoice processing - from Information extraction to client's ERP update. Used Google OCR and NLP (LSTM, Regex). **Saved \$3000 in labor cost monthly**
- **Power BI chatbot**: A POC for Heinekin- allowing upper management to directly query their BI dashboard through natural language. Used Azure Bot Service. **POC was accepted by Heineken and project was financed in Nov.**

RESEARCH INTERESTS

Natural Language Processing and Applied AI/ML • Application Areas: Healthcare and Cognitive Science

SELECTED COURSEWORK

Math Linear Algebra • Calculus • Statistical Inference • Intro to Proofs • Networks

Econ Advanced Econometrics • Statistics for Economics • Microeconomics • Macroeconomics • Game Theory

CS Advanced ML • Advanced Programming • Algorithms • Data Mining & Warehousing

Psychology Intro to Psychology • Thinking like a Psychologist

Research Independent Study: Economics and Computation • Independent Study: Financial Data Analysis

PROJECTS

Twitterazzi | NLP - NER, SENTIMENT ANALYSIS

- Web app that gives a quick overview of an influencer's twitter activity.
- Identifies keywords, entities (Person, organizations & locations) and user's sentiment by processing recent tweets using Spacy, NLTK and Regex.

Reconstruction of Obfuscated Images | ML - CONVOLUTIONAL AUTOENCODERS

- Trained Convolutional Autoencoders on about 13000 images to fix 3 types of Obfuscations(blur, pixelation and speckle noise) in facial images.

Book Spine Detection & Extraction | IMAGE PROCESSING

- Web app that recognizes books in a book shelf image. Segments images by identifying book spines using canny edge detection and Hough transformation.
- Digitizes and extracts the text from the segmented images to identify books.

Generating Synthetic Population | ML - DATASYNTHESIZER

- Worked under **Dr. Debayan Gupta** for studying the spread of Covid-19. Trained a DataSynthesizer model using NSS data to scale the dataset 15 times to about 15 million observations.
- Presented an Iterative Proportional fitting(IPF) approach to integrate census datasets for generating Synthetic Population.

Auto Mobile Slowdown: A Quantitative Study | OLS REGRESSION

- Semester long project under **Dr. Debayan Gupta** studying the Automobile slowdown of India in 2019.
- Performed OLS regression to quantify the impact factors like regulation changes, NBFC crisis, rise in ride hailing services and consumer confidence played.

Hand Gesture Recognition | ML - CONVOLUTIONAL NEURAL NETWORKS

- Trained a convolutional neural network on 1100 images to recognize 11 types of hand gestures. Accuracy $\approx 95\%$

Stocksnapshot | WEB SCRAPING, UI DEVELOPMENT, CLOUD HOSTING

- Developed a stock analysis web app that scrapes over 20 yr+ financial data. Over **100** monthly active users.
- Has a built in DCF calculator for a quick intrinsic value calculation of a stock. Hosted the app on Heroku.

OTHER ACTIVITIES

Well Being Team | FRACTAL ANALYTICS

May 2022 – Present | Gurugram, India

- Helped start a helpline for confidential 1-1 professional counselling. Over **700** sessions booked till now.
- Assisted in conducting anonymous mental health surveys and in organizing relevant webinars.

Investments Club | ASHOKA UNIVERSITY

Sept 2017 – March 2019 | Sonipat, India

- Gave lessons on valuing assets through Discount Cash Flow and Balance sheet analysis.
- Helped start an investment fund under the name of Bodhi Capital, India's first student led fund.

University Sports Team | ASHOKA UNIVERSITY

August 2016 – March 2019 | Sonipat, India

- Swimming - Silver & Bronze in intra-university tournaments. IPSC nationals finalist.
- Shooting - Gold & Silver in intra-university tournament. Represented my college at All India inter-university championship.

SKILLS

Languages

Natural Language Processing

Machine Learning

Data Visualization

Cloud Technologies

Project Management

Python • SQL • Stata • \LaTeX •R • Bash • Java • Django

Regex • NLTK • TF-IDF • Spacy • BERT • LSTM • LDA

Keras • Tensorflow • Scikit-learn

Power BI • Tableau • Matplotlib • Plotly

Azure • AWS • GCP

Git • JIRA • Bitbucket

CERTIFICATIONS

Azure Fundamentals - Microsoft

Google Cloud Fundamentals - Coursera

Math for Machine Learning: Linear Algebra - Coursera

Math for Machine Learning: Multivariate Calculus - Coursera