

# Ajit Kumar Singh

Data Scientist

🌐 [github.com/ajitsingh98](https://github.com/ajitsingh98) | 📁 Portfolio | 🌐 [linkedin.com/sajit9285](https://www.linkedin.com/sajit9285) | ✉️ [sajit9285@gmail.com](mailto:sajit9285@gmail.com) | 📞 +91-8875696158

## EDUCATION

**Indian Institute of Technology, Guwahati**  
Bachelor of Technology in Mechanical Engineering

Jul 2017 - Jun 2021  
GPA: 8.18/10.0

## EXPERIENCE

**HiLabs Inc**  
Data Scientist II

Pune, India  
Jan 2023 - Present

- Revamped the identification of discrepancies between clinical and claims data to enhance **HEDIS scores** in health plans.
- Developed a robust NLP pipeline utilizing pre-trained **SpaCy NER models**, such as **Med7**, to extract diagnosis and procedure codes from unstructured clinical data.
- Leveraged statistical techniques, including **association mining** and **Bayesian belief networks(PGMs)**, to establish **causal relationships** among diverse code sets. This initiative resulted in a remarkable **75%** reduction in NLP pipeline overhead.

Data Scientist I

June 2021 - Dec 2022

- Spearheaded the technical development of **VBCMcheck**, a product focused on detecting data anomalies, and performing root cause analysis of the issues in the Value-Based Care (VBC) domain through **predictive modeling**.
- Leveraged advanced modeling methodologies, including multivariate **time series analysis and forecasting**, incorporating **ARIMA** models and **LSTM-CNN-based** techniques to enhance anomaly detection.
- Built and deployed an end-to-end data ingestion pipeline in the **AWS environment** using the **Apache Spark** framework. This pipeline efficiently processes data from approximately **60** source tables, managing **billions** of records daily.

Data Scientist Intern

Feb 2021 - April 2021

- Developed an automated **NLP-based** data comparison tool as part of the WGS-NASCO **data migration** project.
- Worked on modeling **FP-Growth** based association mining techniques to identify and report potential anomalies.

**Gemini Solutions Pvt. Ltd.**

Gurugram, India

Deep Learning Engineer Intern

May 2020 - July 2020

- Developed a deep learning based **chatbot** for efficient information retrieval from 10-K financial filings of the companies.
- Utilized classical NLP techniques, including **BOW**, **Tf-idf**, and **word2vec** for feature extraction from textual data.
- Fine-tuned a **seq2seq BERT-based** model to generate **contextual embeddings**, enhancing the chatbot's ability to provide accurate and context-aware responses.

## PROJECTS

**MLOps Specialization - Deploying ML Models in Production**

course-certificate

- Deployed ResNet50 and ResNet101 as production and canary models on **Google Kubernetes Engine** using weighted load balancing and request header-based traffic splitting strategies.
- Configured Horizontal Pod Autoscaler and monitored serving performance and resource utilization using **locust**.
- Automated Machine Learning **CI/CD** pipeline using Github Actions for efficient and consistent delivery.

**Semantic Image Segmentation and Activity Recognition**

semantic-segmentation

- Fine-tuned a pre-trained **U-Net PyTorch** model to perform semantic image segmentation, specializing in cooking-related objects and activities within video frames.
- Employed **ResNet** model to enable activity recognition in cooking videos.

**Indigenous Cattle Breed Identifier WebApp**

cattle-breed-classifier

- Built and deployed an end-to-end multi-class classifier based on **residual CNNs**, achieving an accuracy score of around **90%** through techniques such as **hyper-parameter tuning**, **data augmentation**, and **transfer learning**.
- It empowers farmers to accurately identify their cattle's breed based on uploaded cattle images.

**DeepLearning.AI Projects [Coursera MOOC]**

course-certificate

- Implemented cutting-edge **Computer Vision** techniques such as YOLO-based car detection, Neural Style Transfer, and Face verification and detection. Additionally, Worked on creating **NLP-based** algorithms, including LSTM-based jazz improvisation and trigger word detection.

## CORE COMPETENCIES/TECHNICAL SKILLS

**Programming Languages:** Scala, C++, Python

**Big Data Stacks:** Apache Spark, Hadoop, Apache Solr, Spark SQL, PySpark

**Tools/Frameworks:** PyTorch, Tensorflow, Scikit-Learn, NumPy, Pandas, Beautiful Soup, Matplotlib, OpenCV

**Statistics/ML:** Association Mining, Regression, XGBoost, Random Forest, SVM, Transformers, RNNs, CNNs, GANs

**Miscellaneous:** Amazon Web Services (AWS), Git, SQL, Bash Scripting

## RELEVANT COURSEWORK

**Computer Science:** CS101 (Introduction to Computing), CS110 (Lab), Data Structures and Algorithms (GFG)

**Machine Learning:** Deep Learning & Machine Learning Specialization (DeepLearning.AI), Big Data Analytics (Great Learning)

**Mathematics:** Calculus, Linear Algebra, Statistics & Probability, Predictive and Probabilistic Modeling