

INDRAJIT SINGH

LEAD DATA SCIENTIST

"...you realize that all along there was something tremendous within you, and you did not know it." - Paramahansa Yogananda

CONTACT



+91 9591182659



indrajitsingh37@gmail.com



[NLPBlog](#)



Pune, MH-411057



[Linkedin_Profile](#)



[Github](#)

TOP SKILLS

Compassion

Natural Language Processing (NLP)

Data Science

Machine Learning

Programming Languages (Python, SQL etc.)

Data Analysis & Visualization

Kafka, Hadoop & Distributed Computing

EDUCATION

M.Tech (Data Analytics)

Birla Institute of Technology

2016-2018

Dissertation: Recurrent Neural Tensor Network for Predictive Escalation

B. Tech (Information Technology)

West Bengal University of Technology

2008-2012

Studied subjects such as Data Structure, C++, and Automata.

PROFILE

11 Years

Experienced Data Engineer and Software Engineer with a decade of hands-on expertise in data engineering and data science. Dedicated to driving team success through efficient data processing, advanced analytical techniques, and exceptional organizational abilities. Proficient in Natural Language Processing (NLP) and Deep Learning methodologies, applying them to extract insights from large datasets. A seasoned veteran in data engineering, highly motivated to continuously learn, advance, and excel in the ever-evolving field of technology.

WORK EXPERIENCE

Lead Data Scientist

eCelrx, Pune, Maharashtra

May-2021-Present

- VoiceAi and TextAi Solutions Lead (Subject Matter Expert):
- Led development and implementation of VoiceAi and TextAi solutions, leveraging NLP and advanced data engineering techniques.
- Created an automated NLP system, reducing Quality Management processing time by 70%.
- Implemented end-to-end NLP pipeline for Speech+LLM based emotion detection, boosting accuracy by 50% and increasing insurance revenue by 5%.
- Improved speech-to-text model accuracy from 67% to 89% via enhanced data auditing and training.
- Optimized AutoQa code, reducing processing time from 1.5 minutes to 5 seconds.
- Demonstrated Adverse Event Detection POC, securing a \$1M deal in Pharma domain.
- Developed Grammar Analyzer, increasing chat audit capacity by 400%.
- Managed team of 6+ Data Scientists/Analysts for AutoQa, Speech, and VoiceAi.
- Published whitepaper on "NLP for Business-Driven Applications."
- Provided expert consultations for 30+ projects, addressing diverse business challenges.

Senior CSE

Fivetran, Bangalore, KA

2020-2021

- Successfully implemented on-premises ETL knowledge into cloud-based ETL/ELT tools, such as Fivetran, facilitating seamless data integration and transformation.
- Managed and optimized very large data warehouses and data lakes deployed on cloud platforms, including BigQuery, AWS Redshift, and PostgreSQL, ensuring efficient data storage and retrieval.
- Implemented a Quality-driven approach, focusing on three key principles: Understand, Empathize, and Solve, resulting in enhanced customer satisfaction and improved service delivery.
- Attained a perfect 100% customer satisfaction score for consecutive quarters, reflecting exceptional performance and dedication to meeting customer needs.
- Consistently achieved a 99 % Service Level Agreement (SLA) for resolving bugs.
- Built robust data pipelines on cloud platforms, leveraging tools such as dbt, data flow, and model data warehousing, facilitating successful migration from on-premises to cloud environments for enterprise customers.

NLP PROJECTS

VoiceAi- Speech Analytics Solution

Achieved notable success in VoiceAi project, utilizing speech analytics to enhance customer interactions. Proficient in gathering, analyzing, and interpreting large speech data volumes, and designing deep learning models for valuable insights.

AutoQa- Text Analytics Solution

AutoQA enhances customer service quality through automated assurance. It analyzes agent behavior, identifies improvements, and provides actionable insights for better customer experience. This reduces manual costs while delivering accurate and consistent results, saving time and money.

Project Self-Realization

My personal project : An innovative AI solution guiding individuals in discovering life's purpose. Leveraging advanced AI techniques, it provides tools to explore inner selves and answer fundamental existential questions.

CERTIFICATIONS



CellStrat Certified NLP expert

AWARDS

- Super Achiever Award 2023
- The Motivator Award 2022 Q4
- 2022 Bloggathon winner at eClerx
- 3 consecutive appearance in organisation quarterly magazine.

LANGUAGES

English

Hindi

Bengali

WORK EXPERIENCE

Senior Software Engineer

Informatica, Bangalore, KA 2018-2020

- Collaborated with a top-performing team of ETL Engineers and product specialists, significantly enhancing organisational success through streamlined data processes and advanced data pipelines available at that time.
- Transformed business propositions into quantitative queries, effectively collecting and cleaning data for accurate analysis.
- Architected and developed scalable databases utilising SQL and Spark, enabling seamless ETL processes and efficient data management.
- Evaluated and optimised data pipelines, achieving enhanced workflow and increased efficiency for processing over 50 TB of data daily.

Software Engineer

Informatica, Bangalore, KA 2012-2018

- Demonstrated expertise in enterprise data management, building end-to-end pipelines, APIs, and data warehouses for large banks and DW systems.
- Successfully designed and implemented scalable data architectures, ensuring efficient data processing and storage for high-volume data sets.
- Extensive experience with Spark and other big data technologies, proficient in both streaming and batch processing methodologies.
- Developed ETL tools as a software engineer at Informatica gaining valuable insights into data processing and transformation workflows.
- Acquired comprehensive knowledge of Linux/Unix and Windows operating systems, effectively deploying and maintaining enterprise applications across diverse environments.
- Proficient in working with various databases, including MySQL, MS-SQL, Oracle, MongoDB, and AWS Redshift, enabling seamless integration and efficient data retrieval.
- Implemented multi-stage data pipelines utilizing both NoSQL and SQL large data stores, effectively resolving complex customer issues.

NLP Researcher (Unpaid Professional Learning)

Cellstrat 2018-2020

- Extensively built state-of-the-art models using PyTorch, Huggingface, Keras, and TensorFlow, enhancing data analysis and predictive capabilities.
- Successfully implemented multiple projects utilizing large language models (LLMs), showcasing expertise in cutting-edge natural language processing (NLP) techniques.
- Demonstrated excellence in developing and deploying advanced question answering systems, text summarizers, and text-to-speech (TTS) solutions on AWS, GCP platforms.
- Achieved exceptional accuracy of over 93% in high-performance QA models by leveraging ALBERT and RoBERTa-based models.
- Published a research paper titled "BERT - Algorithm Behind Better Search Engines" under CellStrat, contributing to the industry's knowledge and advancements in search algorithms.
- Developed Abstract Text Summarization using LSTM, enabling efficient extraction of key information from large volumes of text.
- Implemented Named Entity Recognition (NER) on CRM log data using CRFsuite and LSTM, enhancing data categorization and analysis capabilities.
- Provided mentorship and guidance to multiple Data Scientists at CellStrat, fostering a collaborative learning environment.