Sidharth Kumar

LinkedIn Github Kaggle Credly

Ph: +91 – 9902785537, 9470703519
Email: Sidharth05kumar@gmail.com
📍 Hyderabad, Telangana, India

Data science professional with 6+ years of experience. Supervising a team as well as building products with focus on alternative datasets using latest technologies. Helping teams to pivot from legacy systems though AI and automations. Also has experience in US healthcare domain. Proven academic records and demonstrated analytical capabilities on competitive platforms.

Education

Degree	Institute	Year	Grade
PGD in Business Analytics (PGDBA)	IIM Calcutta, ISI Kolkata, IIT Kharagpur	2016-2018	8.74
Bachelor of Technology (B. Tech)	NIT Patna	2011-2015	8.80
Class XI – XII (CBSE)	Darbhanga Public School	2009-2011	91%
Class X (CBSE)	Sunshine Prep/High School	2009	90.5%

Key Projects / Experience — GlobalData Plc — July 2019 - Present

- **Multi-Task Private LLM:** Developed consumer grade generative AI capabilities by fine-tuning LLM for various operational task like summarization, tagging, data extraction. Ensured private usage of LLMs by for data privacy and safety.
- **Sense Hub:** Developed a report writing bot. Built a semantic search engine for retrieval and used GPT for Q/A. Created multi-layered chat & semantic engine using agent bots to write quality reports using own dataset. Added existing analysis frameworks.
- **Trend Insights**: Built a daily newsletter and dashboard for real time trend identification from social data along with alert mechanism. Achieved full automation of the process using mail chimp, elastic search, g-sheets, data studio, email APIs.
- **Patents Company Name Normalisation:** Created an algorithm to map patent assignee name variants with an existing company naming system. Utilized various patent & company features and several textual metrics for developing the matching algorithm.
- **Consumer Review Insights**: Created an automated cluster and topic generation for driving insights from drug reviews. Used transformers as an input to density based hierarchical clustering and zero-shot similarity metric for topic selection.
- **ML based Filters:** Created various machine learning based filters on data streams for various business use cases. Built an Innovation Identifier to classify an article into innovation by fine tuning BERT embedding. Developed a model to identify human accounts using the profile images CNN model to help in process of influencer identification.
- **Operations Automation:** Created Web scraping scripts to collect fillings data from 10+ countries using selenium. Automated PPTs for sales using various data sources using python to save manual time and efforts.
- **Analytical Dashboards:** Developed Verdict media websites traffic monitoring for sales team using Google Big Query and Data Studio. Built a competitive intelligence dashboard for company performance benchmarking using Plotly dash framework.
- **Recognitions & Activities:** Best Performer - Business Fundamentals 2021, Best Employee – Q4 2021, CSR Run for Girl Child

Key Projects / Experience — UnitedHealth Group — Oct 2017 – June 2019

- **Retention Driver Analysis:** Built logistic, random forest and gradient boosting models for identifying factors influencing member retention during annual enrolment period. Used Shapley value and partial dependent plots for quantifying the impact.
- **CAHPS Project:** Built ML models for enhancing NPS & Medicare Plan Star ratings by utilising the member survey data. Analysed contract & plans and its likelihood to reach 4-star rating. Predicted members who are likely to provide low scores in survey.
- **Internship (6 months) –** Improved provider side experience and reduced cost by encouraging digital adoption using big data platforms for data analysis. Created a complete automation of ETL process. Received Bravos awards & full-time offer.
- **OPTUM - Data Science Challenge -** 1st Runner up world-wide (1st in India) for wining data science challenge among employees. Built a PCP classification model that can reduce cost by recommending right PCP for members.

Key Projects / Experience — Accenture — Jan 2016 – May 2016

- Developed test cases and scripts to make web applications accessibility compliant for US-based software client. Received appreciations from the manager for the work done.

Academics Projects 2016 – 2018	Kaggle Projects and Achievements 2017 – 2020	
• **Retrieval based chatbot** using Dual Encoder LSTM (RNN) model with a recall of 88% on publicly available Ubuntu Dialogue Corpus	Highest Rank achieved **510** among 100k (**top 1%**) data scientist on the platform	
• **Stock Market Movement Prediction** using social discussions with an accuracy of 60% without any financial feature in the model	**Problem**	**Rank (Tier)**
	Sberbank Russian Housing Market Price Prediction	**56 / 3274 (Silver)**
	Data Science Bowl – Lung Cancer Prediction	**190 / 1972 (Bronze)**
• **Research on Robust Regression** for small dataset. Implemented Campbell's method and simulated using Monte Carlo Simulation	Instant Gratification – Kaggle Internal competition	**88 / 1832 (Silver)**
	NOAA Fisheries – Sea Lions Count	**86 / 385 (Bronze)**
	Zillow Housing – Zestimate Prediction	**240 / 3775 (Bronze)**
	IEEE Fraud Detection in Transactions	**328 / 6381 (Bronze)**

Co-curricular and Extra-Curricular Activities

- **1st** runner-up in stock market event and **2nd** runner-up in arty sticks, an architecture event, conducted at IIT – Patna in 2013
- Participated in several robotics event & workshops at IIT Guwahati (2012), ISM Dhanbad (2011) & NIT Patna in 2012
- Ranked in **top 2%** among 5 lakhs candidates in IIT-JEE in 2012 and **top 2%** among 11 lakhs candidates in AIEEE 2011
- Rewarded a certificate of merit for excellent performance in All India Talent Search Examination (AITSE) held in 2003

Programming	Tools	Data Science & ML/DL Frameworks	Certifications
• Python • R • SQL	Hive, Google-Big Query, Data Studio, Tableau PM- Jira, Trello	Scikit-learn, Numpy, Pandas, Plotly, Dash (Flask), NLTK, SpaCy, Tranformers, Selenium, PyTorch, OpenAI, LLM	• IBM Data Science Professional • Executive Data Science – JHU • Investment Management and ML with Python - EDHEC