Sidharth Kumar

LinkedIn   Github   Kaggle   Credly

Ph: +91 – 9902785537, 9470703519
Email: Sidharth05kumar@gmail.com

Data Science Professional with 5+ years of experience. Currently working in B2B segment with focus on alternative data. Also has experience in US healthcare domain. Proven academic records and demonstrated analytical capabilities on competitive platform.

2008-09	2010-11	2011-15	2016	2016-18	2017-19	2019-Present
Class X CBSE Score: 90.5 %	Class XI-XII CBSE Score: 91 %	B.Tech - NIT Patna Grade: 8.80	Accenture, Software Associate (5 months)	PGDBA - IIM Calcutta, IIT Kharagpur, ISI Kolkata Grade: 8.74	UnitedHealth Group, Internship & Full time Data Scientist (6+14 months)	GlobalData Plc, Senior Data Scientist (36+ months)

Key Projects / Experience	GlobalData Plc	July 2019 - Present

- **Patents Company Name Normalisation:** Created an algorithm to map patent assignee name variants with an existing company naming system. Utilised various patent & company features and several textual metrics for developing the matching algorithm.
- **Data Filter Development:** Created various Machine Learning based filters on data streams for various business use cases. Built an **Innovation Identifier** to classify an article into innovation or not by fine tuning BERT embedding for classification task. Also built a second stage model to summarise the article body. Developed an **Influencer Recognition** model to identify human accounts using the profile images from twitter using a Convolutional Neural Network (CNN) model.
- **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.
- **PR Reports and Automations:** Created a report on Vegan based on social data using predefined business text and keyword searches. Created an interactive dashboard on beers to show case capabilities of utilising blogs data. POC on automation of several legacy one pager reports and email alerts for social trends using python to save manual time and efforts.
- **Analytical Dashboards:** Developed Verdict Media websites traffic monitoring for sales team using Google BigQuery and Data Studio. Created a dashboard using social indicators to analyse the effect of Covid-19 on companies and sectors. Built a Competitive Intelligence Dashboard for company performance benchmarking using Plotly dash framework. Created a UI for Analytics Hub dashboard for analysis & metric generation from alternative data.
- **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:** Worked on finding key drivers influencing member retention during annual enrolment period. Built logistic, random forest and gradient boosting model for identifying key driving factors. Used the concept of shapley value and partial dependent plots for quantifying the impact of key drivers on retention.
- **CAHPS Project:** Worked for enhancing NPS & Medicare Plan Star ratings by utilising the member survey data. Analysed contract & plans star rating and likelihood to reach 4-star rating using logistic classifier. Predicted members who are likely to provide low scores in survey using random forest model.
- **Internship (6 months) -** Worked for Provider side of the business to improve experience and reduce cost by digitisation using big data platforms for data analysis. Created a complete automation of ETL process. Received **Bravos** awards twice and was graded excellent in internship evaluation. Also received pre-placement offer to serve as full time data scientist.
- **OPTUM - Data Science Challenge -** 1st Runner up world-wide (1st in India) for wining data science challenge among employees. Built a PCP recommendation model that can reduce cost by choosing right PCP for members. Final model was built using lots of feature engineering and parameter tuning. Achieved an AUC of **86.66** on private leader board.

Key Projects / Experience	Accenture	Jan 2016 – May 2016

- Worked for US-based software client to make web applications accessibility compliant. Involved in preparation of test cases & scripts, running of test scripts, and logging bugs. Received appreciations from the manager for the work done.

Academics Projects	2016 – 2018	Kaggle Projects and Achievements	2017 – 2020

Academics Projects (2016 – 2018)

- **Retrieval based chatbot** using Dual Encoder LSTM (RNN) model with a recall of 88% on publicly available Ubuntu Dialogue Corpus
- **Stock Market Movement Prediction** using social discussions with an accuracy of 60% without any financial feature in the model
- **Research on Robust Regression** for small dataset. Implemented Campbell's method and simulated using Monte Carlo Simulation

Kaggle Projects and Achievements (2017 – 2020)

Highest Rank achieved **510** among 100k (**top 1**%) data scientist on the platform

Problem	Rank (Tier)
Sberbank Russian Housing Market Price Prediction	56 / 3274 (Silver)
Data Science Bowl – Lung Cancer Prediction	190 / 1972 (Bronze)
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

- Secured **1**st runner-up position in StockMart, a live stockmarket event, conducted under Anwesha, IIT – Patna in 2013
- Secured **2**nd runner-up position in Artysticks, an architecture event, conducted under Anwesha in 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	Hadoop – Hive Google-Big Query Google Data Studio	Scikit-learn, Scipy, Numpy, Pandas, Plotly, Dash, NLTK, H2O, Keras, SpaCy, Tranformers	IBM Data Science Professional Executive Data Science – JHU Machine Learning – UL