Daksh Agarwal

Dehradun, Uttarakhand, 248001, IN

(+91) 9418555015 / 7973045391 • ag.daksh@gmail.com
https://in.linkedin.com/in/dakshagarwal

Business-minded Data Scientist with almost 5 years experience in building data-intensive applications, and having the passion to solve real-world business challenges using data analytics. Proficient in deploying complex machine learning and statistical modeling algorithms/techniques. Highly skilled in machine learning, data visualization, and creative thinking.

Experience

- **National Payments Corporation of India (NPCI)** — **Hyderabad, Telangana, IN**
  Senior Associate - Data Science (Market Innovation) — *July 2018 – Present*

  (*National Payments Corporation of India (NPCI) is an umbrella organization for all retail payments system in India. It was set up with the guidance and support of the Reserve Bank of India (RBI) and Indian Banks' Association (IBA).*)

  - Developing various end to end AI applications tailored for specific use-case using state of the art machine learning, deep learning, natural language processing with a focus on business impact and need.
  - Using AI and ML techniques to detect the fraud in highly unbalanced data of the financial transactions across various products of NPCI using Python, Pyspark and other Big data technologies.
  - Interacting with the business people for understanding their business problems and formulating them to be solved as Analytics (Statistical/Machine Learning) problem.
  - Social media analysis to know the sentiments of people towards various products of NPCI.
  - Building web-service API's for internal and external consumption.
  - Creating real-time data pipelines using Apache Nifi, Kafka, Hive, etc.
- **Spineor Webservices Pvt. Ltd.** — **Mohali, Punjab, IN**
  Data Scientist — *Nov 2016 – April 2018*
  - Developing a framework for generating the recommendations for the auto parts warehouse portal.
  - Extract requisite information from the data collected from various clients.

Education and Awards

Academic Qualifications

- **University of Hyderabad** — **Hyderabad, Telangana, IN**
  M.Tech (Artificial Intelligence) , Cum. GPA: 7.97/10 — *July 2014 – June 2016*
- **UIIT, H.P. University** — **Shimla, H.P, IN**
  B.Tech (Information Technology) , Percentage: 72.8% — *July 2009 – June 2013*
- **Govt. (Boys) Shamsher Sr. Sec. School** — **Nahan, H.P, IN**
  Intermediate, Percentage: 71% — *March 2009*
- **Mount Carmel Sr. Sec. School** — **Palampur, H.P, IN**
  Matriculation, Percentage: 81.57% — *March 2007*

Awards and Recognition

- Our team (three members) won the digipay hackathon in the 'Innovation in digital payments' category, organized by Thoughtworks in collaboration with NPCI in April, 2020.
- Received 'Significantly above expectations at work' award for 2018-19 and 2019-20 fiscal years at NPCI.
- Won the 'Key Contributor Award' for the 'Recommendation Systems' project completed in Nov, 2017.

Projects

- **'Fraud detection in RuPay cards'** | *NPCI*

  Worked on detecting the frauds happening with the usage of RuPay cards (ECOM). Achieved the FP rate of 1:20 for few customers, thereby detecting 80 percent of the frauds. Worked end-to-end to execute, implement, and deploy this model at scale with response time for each transaction as <20ms.
  Technologies used: Pyspark, Hive, Python, Redis, MongoDB

- **'UPI (Unified Payments Interface) fraud detection'** | *NPCI*

  Worked on detecting the fraud in highly imbalanced (almost 8k fraud cases in 30 billion records) data of UPI (P2P). Achieved FP rate of 1:40 for some customers, thereby hitting the accuracy of 60 percent. Working end to end, brought this model to production, thereby declining fraud transactions in real time.
  Technologies used: Pyspark, Hive, Python, Redis, MongoDB

- **'OCR for bill-invoices: Obtaining text from scanned documents'** | *NPCI*

  Dataset consisted of product-wise (AEPS/NFS/UPI etc.) bill-invoices (in form of pdf's) of 35+ banks, each bank following a different format. The aim was to fetch some common elements like name of the bank, customer name, and the amount mentioned in the bill invoices. The output consisted of a bank-wise csv file having all the required information for different bills.
  Technologies used: Python, Library: pdfminer, camelot, excalibur

- **'Conversational AI: Chatbot'** | *NPCI*

  A chatbot was developed using RASA, which is an open source library for NLU and ML-based dialogue management. This chatbot was capable of answering most of the queries related to NPCI products and also, it could give you the current transaction status of any transaction related to UPI.
  Technologies used: RASA, Python, NLP

- **'Sentiment Analysis on Social Media Data'** | *NPCI*

  Sentiment analysis was done for the various products of NPCI like Rupay, BHIM, IMPS, BBPS, etc., on the data extracted from the Twitter handles and various other sources performing web crawling and using some tools. Data was parsed and was categorized into three classes, viz. positive, negative, and neutral. Further, the negative tweets were sub-categorized into various buckets like 'Transactions issue', 'App problems' etc.
  Technologies used: Python, Radian6, NLP

- **'Recommendation System for an Auto-parts Warehouse Store'** | *Spineor*

  Implemented a hybrid model (item-to-item and user based recommendation engine) using the Tensorflow's wide and deep model. Ratings for given products were generated which were used to populate the 'user to product' matrix. Cosine similarity between the products/users were used to generate and rank recommendations.
  Technologies used: dl4j, Scala, Spark, DSE Search, Cassandra

- **'A Framework for Navigation in Outdoor Environment'** | *M.Tech Project*

  Developed a framework that can help a blind person in safe navigation in outdoor environment. It employed video stabilization, object tracking with their categorization, feature selection, and recognition of objects in the video.
  Technologies used: Matlab

Technical Skills

Key Skills:

Data Analysis | Data Visualization | Machine Learning | Statistical Analytics | Sentiment Analysis | NLP | Pyspark | Predictive Modeling and Analysis | Model Deployment | Elastic Search | Hadoop Platform | Python

Industry Software Skills/ Tools:

LaTeX • PyCharm • RStudio • Zeppelin • Jupyter Notebook • Power BI • Radian6

Databases exposure:

• MySql • Cassandra • Hive • Redis • MongoDB