

Akash Chauhan

Senior Data Scientist

Jaipur, Rajasthan

📞 (+91) 8233186472 | ✉️ chauhanakash2303@gmail.com | 🌐 chauhanakash23.github.io | 🔗 chauhanakash23 | 📄 2303akash-chauhan

Accomplished Senior Data Scientist with 6+ years of experience specializing in advanced recommendation systems and Natural Language Processing. Demonstrated expertise in designing and deploying high-impact machine learning models, consistently driving improvements in business metrics. Adept at leveraging data-driven insights to inform strategic decision-making and deliver scalable, revenue-generating solutions. Lately been passionate about exploring the potential of Generative AI to innovate and enhance business strategies.

Work Experience

Pocket FM

Remote

Senior Data Scientist

Aug. 2022 - Present

- **Recommendations:**
 - **Enhanced Personalization Engine:** Upgraded the core ranker from a basic ALS model to a sophisticated EASE model, boosting recommendation accuracy and cutting computational costs.
 - **Hybrid Ranker Development:** Engineered a cutting-edge LightGBM-based ranker that leverages the EASE model as an input feature along with hundreds of extensively curated features, optimizing user engagement by predicting x minutes of play on shows, which powers key features like feed and autoplay. This core ranker increased 20% playtime overall from autoplay and the feed.
 - **Strategic Improvements:** Currently integrating revenue optimization into the recommendation system, aligning engagement metrics with revenue goals for comprehensive business impact.
- **Feed Personalization:** Leveraged the core ranker model to design dynamic modules tailored to users' tastes, incorporating sub-genres, tags, and tropes, as well as popularity and trend-based elements. Developed an XGBoost model to personalize the order of these modules, resulting in a cleaner, more dynamic, and relevant feed (as validated by improved NDCG, MRR metrics offline and engagement metrics in A/B testing). This enhancement led to a 19% increase in new show explorations and improved long-term retention (25+ hours) on shows explored from the feed by approximately 15%.
- **Pricing:**
 - Led the development of product strategies like discounting, custom packs, and binge passes, supported by XGBoost, CatBoost, and LightGBM models. These strategies, rolled out after successful A/B testing, improved conversion by 50% and reduced spender churn by 30%.
 - Currently testing a subscription model that predicts user spending capability after 30 days, based on interactions during the initial 3 days, to refine pricing strategies.
- **Growth:** Collaborated with the product team to implement data-driven strategies, successfully re-engaging 30% of churned users with personalized notifications and FOMO-based incentives, leading to a 10% increase in user engagement.

Demandbase

Remote

Data Scientist

Apr. 2021 - Aug. 2022

- **Innovated Data Products:** Spearheaded the creation and continuous enhancement of data products, including renewals and fleet estimation (Regression models), serving 10+ high-profile accounts, driving significant business value.
- **Boosted Customer Acquisition:** Enhanced offerings such as Premium Intent (propensity modeling) and Next Tech Purchase (Market Basket Association), and developed Trending Skills (Curve fitting) to identify high-value targets, resulting in a 12% increase in up-sell opportunities.
- **Revenue Generation:** Played a pivotal role in driving revenue growth, generating over **300k USD** across multiple deals within a single year through the data products like Next Tech Purchase and Trending Skills.

Straive (Formerly SPi Global)

Chennai

Data Scientist

Jun. 2019 - Apr. 2021

- Developed a question-answering approach to extract Environmental, Social and Governance data from unstructured PDF documents(Annual/CSR reports of companies) to reduce the annotators' time by whooping 80%.
- Developed a suite for a legal use case that included summarization and citation recognition using NER techniques.
- Created a deep learning based indexer for medical research articles/papers against concepts(60k+ concepts).
- Set up an internal text summarization platform(after market research compiled models like PEGASUS, ProphetNet, T5 etc) with a REST api and web interface for internal use.
- Combined results of object detection models with rules and patterns(manually created) to extract data from PDFs(Invoice/Debt notices).

Data Scientist Intern

Noida

Tata Consultancy Services

Mar. 2019 - May. 2019

- Created and streamlined an ETL pipeline to consume data from OMOP CDM and extract and store the incidence and prevalence of diseases.
- Planned and developed non-objective based market segmentation models for pharmaceutical market.

Institutional Research Intern

Kharagpur

SWAN Labs, IIT Kharagpur

May 2018 - Jul. 2018

- Developed a novel learning framework for resource constrained IoT devices leveraging the edge of the network, working with two of the finest research scholars of Prof Sudip Misra.

Education

Central University of Rajasthan

Ajmer, Rajasthan

Integrated M.Sc. in Computer Science with specialization in Artificial Intelligence · 5.44/6

Jul. 2014 - May. 2019

- **Gold medalist** - batch 2014-2019.
- Recipient of merit scholarship for 4 years.

Blogs & Personal Projects

Generating abstractive summaries using Google Brain's PEGASUS.

Jun. 2020

- Since the feature to generate summary for a single piece of text was missing, this article shows a neat work-around to generate summaries using the available pre-trained model.

Training a custom NER model using Flair

Jun. 2020

- A tutorial on how to create an NER model for custom entities using Flair. Right from data preparation to model evaluation.

Personalized Cancer Detection

Feb. 2019

- Leveraging the NLP techniques, build models to successfully predict the class of cancer, given the 'Gene', 'Variation' and the corresponding text in the literature.

Stack Overflow tag predictor

Jan. 2019

- Given the title and body of the question on stack overflow, predicting the possible tag(s).

Certificates

2018 **Microsoft** AI Challenge India

Online

2018 **GIAN** Machine learning

CURaj

2018 **GIAN** Big data analysis and mining in deep web repositories

CURaj

Skills

- Intelligence Gathering
- Analytical and Critical Thinking
- Problem Solving
- Machine Learning
- Exploratory Data Analysis
- Deep Learning
- Natural Language Processing
- AWS and GCP Cloud
- PySpark
- Python
- Feature Engineering
- Tensorflow and Pytorch
- A/B Testing
- ANN CNN RNN
- Clustering
- Databricks
- Visualisation
- Apache Airflow