

Mohd Ghouse

Data Scientist

Pragmatic, Competitive and Efficient Data scientist with more than 3 years of experience in application, development and deployment of Machine learning and Deep learning models in NLP(Natural Language Processing), Image processing, Chabot, retail and legal domain Using state-of-the-Art technologies like Transformer models.

Uncovered insights and trends through data analytics and visualizations using Tableau.

Work Experience

Data Scientist

Hobasa India Pvt Ltd, Hyderabad, India – Jan 2021 - present

- Developed a pipeline for Data (Key Entity) extraction from regulatory Forms and Invoices by exploiting deep learning models (layoutlm, layoutparser and Document question answering).
- Extracted Machine readable rules from legal ECFR text data by combining condition and action extraction along with entities extraction through Flair model.
- Fine tuned Flair Model for entities extraction on ECFR data and added custom entities to the model required by the business.
- Built a pipeline for preparing the data for training the Flair entities extraction model which bridges the gap between data annotation and data ready for model training.
- Developed and deployed a FAQ's Chabot using rasa framework

Data Scientist (Intern)

INSOFE Pvt Ltd, Hyderabad, India – Jan 2020 – Mar 2020 AND Aug 2020 – Oct 2020

- Predicted fail changes from HPSM tool for a client with a trade-off between recall and precision.
- Addressed class imbalance (95:5 ratio).
- Identified patterns in the data employing seaborn and matplotlib libraries of python.
- Assortment planning and Demand forecasting for retail client.
- Uncovered sales pattern of different products through Data Analysis by Tableau.
- Disclosed typical customer behavior using RFM Analysis

RF Engineer

Toshniwal Enterprises Controls Pvt Ltd, Hyderabad, India-- Aug 2016 – Jul 2017

- Reduced call drop rate by almost 20% by optimizing RF sites and tuning signal parameters.
- Increased DL throughput from 30 Mbps to 45 Mbps by identifying and eliminating swap.
- Reduced ping response rate for highly dense clusters by reducing overshooting of neighboring sites.

Achievements

- Enhanced unstructured and semi-structured to structured pipeline by replacing Microsoft computer vision API (paid service) with a transformer model for handwriting recognition and reduced the project expenses by approximately 10%.
- Curtailed research time for stake holders and SME's by 50% to 60% by developing a similarity search model on ECFR corpus.

CONTACT

- HYD, India (Open to Remote)
- +91-9700777073
- ghousesalman9@gmail.com
- linkedin.com/in/ghouse-salman-974656122

SKILLS

Data Visualization:

- Tableau (Advanced)
- Seaborn (Advanced)
- Matplotlib (Advanced)
- DOMO (entry level)

Techniques:

- Document Question Answering
- Entity Extraction from text
- Text Summarization - Transformers
- Similarity Search - Transformers
- Clustering and Topic modeling
- Predictive Modeling
- Data annotation - Prodigy

Tools and Frameworks:

- Python, scikit-learn, pytorch
- Data Analytics – Pandas, Numpy
- Databases (MySQL)
- Docker
- Flask
- Rasa – Chabot
- Azure Databricks

EDUCATION

PGP in Data Science and Optimization

Major in Machine Learning and Artificial Intelligence
Hyderabad, India – 2019

SHADAN college of Engg and Tech
Major in Electronics and Communication - E.C.E
Hyderabad, India - 2015

OTHER

- Stock Price Prediction of Company - (Hackathon Conducted By INSOFE)
- Sentiment Analysis of Lion King Movie Reviews