Rishabh Mehra

E-mail: rishabhmehra5@gmail.com ✻ *Telephone number:* +91-9619851863
GitHub: https://github.com/RishabhMehra
LinkedIn: www.linkedin.com/in/rishabhmehra0512

BIO

I am a Data Scientist with about 4 years of experience in developing machine learning algorithms and data pipelines. My expertise lies in NLP, particularly in creating NER models to extract information from unstructured data. I collaborate with cross-functional teams to identify business problems and develop solutions.

Work experience

I3Systems.ai *March 2022 – Present*
Data Scientist *Mumbai*

- Upgraded **Document and Image classification** models **improving the accuracy score by 5% over existing model**.
- Built Extraction pipeline to work with financial documents such as **Bank Statement, Computation of Income and Form 16** utilizing **NER models, string matching rules, section classification** to extract relevant information.
- Worked with **Azure Form Recognizer** for Table extraction from documents , trained new models and worked using existing models as well.
- Developed an algorithm that extracted salary data from bank statements with **97 % accuracy**.
- Improved **data labeling efficiency** through identification of different document formats and annotation patterns, resulting in more accurate and reliable results in natural language processing, computer vision, and machine learning applications. **Reduced costs and shortened development cycles** while **increasing the speed and effectiveness** of solutions.
- Trained a NER model using **Flair & CRFPP** to recognize important information from different COI formats with a combination **accuracy of 85%.**
- Developed an algorithm and model to analyze documents with a standardized layout. This project involved extracting relevant information from **unstructured data to create a structured outputs**.
- Created MER(Medical Examination Report) **Documents Digitization pipeline** using **Image Localization** techniques to extract data points, which improved data quality and minimized human error.
- Developed a **regression model** to find data point coordinates, which improved data extraction efficiency.
- Developed a **Motor Invoice Digitization pipeline** from scratch, utilizing text and table extraction with standardization techniques. Proactively identified and addressed potential issues, resulting in **consistent high-quality results**.
- Developed an **auto-learning framework** for profile terms with range values reducing compute time by**7 second per case, and reducing manual user intervention by 50 hours** .

Larsen and Toubro Infotech(LTI) *August 2019 – February 2022*
Senior Software Developer *Navi Mumbai*

- Evaluated project requirements and specifications and developed software applications using **K2 Backpearl and K2 five** that surpassed client expectations.
- Gathered and defined customer requirements to develop clear specifications for project plans .

- Analyzed issues report to find problem areas and was able to **reduce ServiceNow Ticket count by 60 %** .
- Utilized exploratory data analysis techniques to identify patterns, relationships, and trends.

Agrima Infotech *October 2018 – January 2019*
Data Science Intern *Remote*

- Developed custom data models and algorithms.
- Provided a quantitative framework for evaluating and analyzing alternative techniques.
- Recommended and implemented changes that streamlined processes and improved usability.

Education

B.Tech - IT *Vidyalanakar Institute of Technology, Mumbai*
July 2015 – April 2019

PSYCHOMETRIC ANALYSIS USING SOFTWARE SIMULATION
Recommending candidates based on recruiter requirements and score obtained in situational judgment test. Applied **word embedding** on resume to recommend profiles. Designed a scoring mechanism that takes into account user preference and profile matching.**Design Paper Published on IJCA** - International Journal of Computer Applications.

Technical skills

Programming Languages/Tools	Machine Learning, Deep Learning, Python, Flask, NLP, Pytorch, SQL, Azure, Tensorflow, Git, Spacy

Projects

Wafer Fault Detection

- To build a classification model to predict the quality of wafer sensors based on training data.
- Develop a pipeline to automate the training and prediction process it included data validation, file validation.
- Used clustering and classification models like **K Means ,Random Forest and XGBoost** for each cluster

Best Tweets for a Query

- It scrapes the tweets from twitter in real time and runs an algorithm over it to display the best tweets. It is implemented end to end and hosted it on Heroku .Developed an algorithm to find the best tweets for a given phrase/word using NLP techniques like **word embedding, sentence embedding**.

Attendance System based on Face Recognition

- Aim : To detect the employees details for attendance purpose.
- Used **MTCNN with Arcface** to develop the face recognition model, Development a UI for the project using Tkinter.
- GitHub link : https://github.com/RishabhMehra/Attendance-System