NASHIT BABBER

SUMMARY

- Business oriented *Full-Stack Data Scientist* with demonstrated ability to mine hidden gems in the ocean of data and extract meaningful insights from it and develope cutting edge production-ready front end for it.
- Data Science, Data Engineering & Data Integration professional with around 6 years of experience working on end to end Machine Learning, Data Engineering, Data Visualization and Data Warehousing Projects.
- Currently working as a Data Scientist in Governance, Risk & Compliance of Wolters Kluwer India.
- *Expert in Python Machine learning stack (Numpy, Pandas, Scikit-Learn, Matplotlib)*
- *Skilled in Machine Learning, Deep Learning, Natural Language Processing and Predictive Modeling and Hadoop*

EXPERIENCE

Data Scientist, 03/2018 to Current
Wolters Kluwer - Pune
Legal Bill View Analyzer

- Collaborated with several domestic as well as international clients for requirement gathering and developing production-ready codes for them.
- Worked on the problem of *Label Correction of Imbalanced* Datasets to increase the AUC of an ML model in production by 10-15%.
- Responsible for building and maintaining *Machine Learning pipelines end-to-end* in Python as per the need of the Insurance clients.
- Built an *ML Model* Recalibration *Framework* that decreases the recalibration time by *60%.*
- Worked on the POC pipeline of Legal View Bill Analyser to analyze and re-fit existing ML pipeline on unseen data for new clients.
- Researched and Developed client-specific Deep NLP models to boost precision or recall or both as per the business requirements.

Technologies Used : *Imblearn, Python, Sci kit Learn, Matplotlib, Regex,doc2vec, LDA(Topic Modeling), Random Forest, XGBoost, Plotly, K Means Clustering, Custom Word Embeddings, TFIDF, Text Rank, JSON, NLP.*
Reasonableness Review Tool

- Single-handedly identified the pain point, procured business requirement, Developed and Deployed production-ready front end and back end python-based dash application for in house reviewers.
- This project saved in house reviewers time by 80% and decreased

CONTACT

Address: Pune India
Phone: +91 - 7222956561
Email: nashit93@gmail.com

SKILLS

Languages
- Python, R,HTML,CSS,JavaScript

DataBases and Cloud Services
- MySQL, JSON, Google Firebase

Tools And Frameworks
- Pyspark ,Hadoop Ecosystem, Docker, Flask, Streamlit, Git ,Dash

Libraries
- Scikit-learn, pandas, numpy, re, keras, tensorflow, huggingface, transformers, gensim, fasttext, nltk, spacy, flask, xgboost, *Selenium , Beautiful Soup(BS4) , lxml* ,plotly and many others

WEBSITES, PORTFOLIOS, PROFILES

https://github.com/nashit93
http://nashit93.github.io/

WHEN NOT DOING MATRIX

human error by 30%

- Got *appreciation mail* from all the end-users for *discovering the pain point* and resolving them and making the interface so user friendly.

Technologies Used : *Python, Dash, Plotly, Docker*

File Splitter Tool

- Doing POC for any new client of LBA was very tedious and prone to error job. There were often a lot of back and forth communications between the reviewers and the data scientists.
- Discovered the *pain point* and created a tool which would automatically perform the POC and generated the output and can be used directly by reviewers with no coding background and experience.
- This helped in the reduction of human errors and miscommunication errors. Also, decreased the time taken for POC by a large factor. *The job that took days to complete is now completed in a matter of hours.*
- Got Appreciation award and mail for the same.

Technologies Used : *Python, StreamLit, Scikit - Learn, XGBoost, Docker*

Legal Document Keywords Extraction

- Rapid prototyping of reading legal documents and summarize them using AI.This saved lawyer's reading time to find new rules from the document and created a keyword extraction algorithm to extract and highlight the key phrases from the document.
- Worked extensively on document tagging , to check if rules in the documents are relevant to the forms or not. Created a dictionary for semantic match of the words between the documents and the forms, Saving a huge amount of lawyer resources time and money for the company.

Technologies Used : *Reg Ex , Elastic Search , Text Rank , K mean Clustering ,Sent2Vec,Rake, distance Algorithms*

Lien Solution

- Collaborated with onsite team members to analyse a firm's debtors so that they can understand the customer behavior and prioritize their lien.
- Played with extremely dirty data, name variation.
- Used Fuzzywuzzy, Levenshtein distance and regular expression to clean the name.

Technologies Used : *Pyspark, john snow labs*

Machine Learning Engineer, 04/2017 to 03/2018
Hezkrost Technologies - Pune, India

- Co - Founded the company.
- Designing and Development of an extremely efficient RPA system in python, which could run on low end computers(Such as Raspberry Pi) and scrap,clean, process stocks data for 1800+ stocks, their daily Highs/Lows , Advances / Declines, Bulk Deals , Quarterly income tax statements , Balance Sheets and Cash Flows.

MULTIPLICATION I LOVE TO

- Swim
- Play Badminton
- Table Tennis
- Computer Games
- Go on long drives

CAUSES

Founder At FreeJee.com - Providing free education to those in need

- Collaborating with other team members to code Highly Efficient HFT algorithms for real time Buy/Sell triggers on live market data of around 1800+ stocks, using various technical algorithms such as SMI, MDA, oscillators, Bollinger, Pivot Points, RSI, super trend etc. *These algorithms were so efficient that that made decision in ~3ms for 1800+ stocks.*
- Create a Stocks news sentiment classifier and Stock recommendation system using state of the art deep learning techniques.
- Integrating news sentiment and recommendation engine's output to HFT algorithms to get more precise stock buy sell triggers which predicted the triggers 25% better than traditional figures.

Technologies Used : *Python , Selenium , Beautiful Soup(BS4) , lxml , pandas ,*
numpy, Keras, Doc2Vec , Gensim.

Associate Consultant, 08/2015 to 04/2017
Saama Technologies - Pune, India

- Rapid Prototyping and demonstration of applications for real world evidence in the pharmaceutical industry.
- Developing and presenting the models for integrating social media data reporting patient's real experience of treatment, including drug efficiency, safety and drug impact. Improved scoring time(by 25%) and accuracy(by 5%) by implementing a blend of rule based system and machine learning.
- Data Migration of fresh and incremental data load from TeraData to HDFS using Pyspark.

Technologies Used : *HDFS , Sqoop , Pyspark , Hive , Impala , Python , Shell Scripting , Machine Learning.*

EDUCATION AND TRAINING

Bachelors of Engineering, Mechanical Engineering, 06/2015
University Of Pune - Pune

ACCOMPLISHMENTS

- Led team to make end to end tools, earning recognition from upper management and financial reward.
- Awarded ‘Star Performer of the Quarter' for the periods Q2-2019 and Q3-2020
- Received ‘Standing Ovation Award' from the EVP for the period Q3-2020.
- Won the company-wide hackathon competition on developing a chatbot within 24 hours using just encoder decoder model
- Delivered training on Python basics and Python for Data Science to 25 company employees.

CERTIFICATIONS

- Completed Deep Learning Specialization, Natural Language Processing, Machine Learning and 6 other courses on Coursera.
- Completed Google Tensorflow certification Exam(Result Awaited)