BAIJU KUMAR SINGH

Data Science Engineer

E-Mail: singhbaiju94@gmail.com
GitHub ID: https://github.com/singhbaiju94
Phone: +91 9106242207

CORE COMPETENCIES

Software Development

Data Science & Analytics

Requirement Gathering

Machine Learning

Natural Language Processing

Java Scripts

OCR

Kafka

PROFILE SUMMARY

- An enthusiastic professional with **4 years** of experience in **Data Scientist** using **Python, NLP & Machine Learning**
- Proficient in **Data Analysis, Data Science and Machine Learning** Operations
- Rich **project management & governance** experience, proficient in grasping the big picture, conceptualizing, developing & implementing solutions, partnering closely with business leaders & stakeholders to achieve higher levels of efficiency and effectiveness
- Analysed organizational performance through **Data Science techniques** to derive organizational KPIs/ metrics and data visualization solutions

ORGANISATIONAL EXPERIENCE

09th MARCH 2021 – Present   United Health Group Bangalore as Data Scientist

Roles and Responsibility:
- Working on Big Medical data to create end to end pipeline for processing and Business Analysis.
- We use python and pyspark language to analyse the big data.
- We are using Apache Spark and Kafka framework to handle big data.

17th June 2017 - 08th MAR   2021 Larsen & Toubro Technology Services Limited, Bangalore as Data Scientist

Roles and Responsibility:
- I have worked on structured and unstructured data from Oil & Gas domain.
- I have done document classification, Information detection and extraction using Python, NLP.
- I have Built Classification model for document classification.

Project Undertaken

Title: HRA Data Mart	Organisation: United Health Group	Team size 2

Role: Developer   **Technology: Pyspark**

Description: To develop end to end big data (Medical data) pipeline for cleaning, processing and Business Analysis.

Responsibilities:
- Read the daily basis data from the hive server and ingesting the HDFS server.
- Using Apache spark Process the huge amount of HDFS data and perform the pre-processing and business analysis.
- After performing business analysis store the data into data ware house.
- Fetch the data from data ware house and pulling the data into business analysis tool Power BI.

Title: Ainfonix	Organisation: L&T	Team Size: 3

Role: **Developer**   **Technology: Python, NLP, Machine Learning**

Description: This is an automatically information detection and extraction utility from the various kinds of engineering documents.

Responsibilities:

- Developed a tool that segregates good documents quality file from bad ones using python and NLP.
- Template Segregation of different kind of files such as reports, datasheet and p&id which will reduce manual efforts using python, NLP, and Machine Learning.
- Write module & applications using Python by leveraging existing organizational NLP and ML based platform to accomplish Data Extraction, Tag Extraction, As-Set register information from the multiple documents to very high precision & recall.
- Hosting and testing in-house developed tool to the client environment.
- Actively involved in communication with client and functional team and ensuring consistent user experience.

Title:	Tag Prediction Model	Organisation: L&T		Team Size:	2
Role:	Developer		Technology:	Python, NLP & Machine Learning	

Description: This is a machine learning based tag classification tool for predicting the similar and correct tag.

Responsibilities:
- Collect the pre executed large amount of dataset and performed processing techniques and create features from different types of tags.
- Based on the feature to train the model for predicting the similar tags using machine learning classification algorithms.
- Conducted training & testing of module; ensured accuracy of the model by cross validation and predict the tag class accordingly.
- Actively involved in communication with client and functional team and ensuring consistent user experience.
- Hosting and testing in-house developed tool to the client environment.

Title:	Customized NER Model	Organisation: L&T		Team Size:	3
Role:	Developer		Technology:	Python, NLP & Machine Learning	

Description: This is a Named Entity Recognition Model that automatically identifies named entities in a text and classifier them into predefined categories.
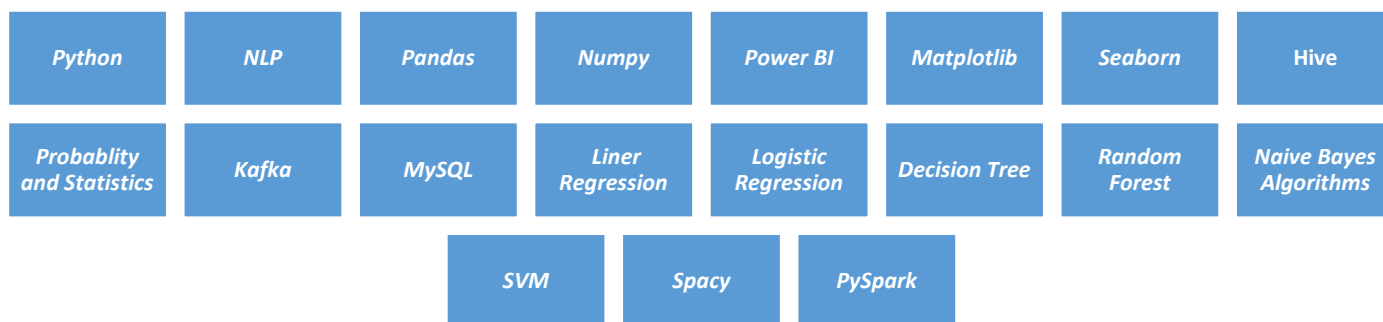
Responsibilities:
- Created a domain dictionary with all inflected forms/variations and their respective values.
- Based on the prepared dataset we created a NER training dataset format as per spacy standard.
- Extracted and classified the named entity attributes from unstructured data.
- Actively involved in communication with client and functional team and ensuring consistent user experience.

EDUCATION

- **BE - (Computer Science)** RGPV University, Bhopal 2016; secured 7.30 CGPA
- **12th** - Bihar School Examination Board, Patna 2011; secured 60.5%
- **10th** - Bihar School Examination Board, Patna 2009; secured 74.5%

TECHNICAL SKILLS

Python	NLP	Pandas	Numpy	Power BI	Matplotlib	Seaborn	Hive
Probablity and Statistics	Kafka	MySQL	Liner Regression	Logistic Regression	Decision Tree	Random Forest	Naive Bayes Algorithms
		SVM	Spacy	PySpark			

PERSONAL DETAILS

Date of Birth:	15th December 1994
Languages Known:	English and Hindi
Kaggle Id:	https://www.kaggle.com/singhbaiju
Linked Id	baiju-kumar-singh-881803115/
Address:	Mjr Pearl Flat no 301, I Tower opposite Whitefield railway station Bangalore