## **BAIJU KUMAR SINGH**

## Data Science | Machine Learning | NLP Developer

E-Mail: singhbaiju94@gmail.com

GitHub ID: https://github.com/singhbaiju94

Phone: +91 7338631508

# **CORE COMPETENCIES**

**Software Development** 

**Data Science & Analytics** 

**Requirement Gathering** 

**Machine Learning** 

**Natural Language Processing** 

**OCR** 

**PySpark** 

**Microsoft Azure** 

**Probability and Statistics** 

## **PROFILE SUMMARY**

- An enthusiastic professional with 3.8 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

### **EDUCATION**

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

## **ORGANISATIONAL EXPERIENCE**

Jun'17 – Present Larsen & Toubro Technology Services Limited, Bangalore as Data Scientist Key Result Areas:

- Managing the activities including designing, developing, testing, troubleshooting and debugging the applications.
- Providing post-implementation, application maintenance and enhancement support to the client with regard to the product/ software application.
- Conducting brainstorming sessions to assess & evaluate applications for new tools & technologies as these continuously evolve.
- Leading the software development activities for business process mapping of the client and identifying appropriate development methodology.
- Enhancing the flexibility of application to ensure its adaptability as per customers; conveying technical functionality to a non-technical audience.
- Monitoring data analysis and processing activities involving analysing, studying and summarizing data for extracting useful information which would assist in strategic decision-making and planning.
- Collating appropriate data for use in database and conducting that the related research.
- Identifying, evaluating and mitigating risks to ensure operational efficiency at all times.
- Evolving critical business solutions through information gathering, synthesis, review, and testing.
- Adding features to the product as per the requirements from client; mapping business requirements and providing customized solutions involving finalization of product specifications and selection of appropriate techniques.
- Executing client/ user/ partner interaction for requirement gathering, risk assessment, finalization of technical specifications and discussions for effort/ cost/ time estimation and client coordination and reporting.

Title: Ainfonix Team Size:

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 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 Team Size:

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.

#### **Projects on Machine Learning Algorithm**

#### **Title: Automobile Price Prediction**

**Description:** To develop a model that can be used to predict the price of a new car based on the market survey.

#### Responsibilities:

- Designed Linear Regression model, Ridge regression model and Lasso Regression model.
- Executed various performance measures including R-square & mean-squared error to select the best model.
- Attended accuracy of 73% with Ridge Regression model and used this model to predict the price of car.

#### **Title: German Credit Risk Prediction**

**Description:** To develop a model that can be used to determine if a new customer is a good credit risk or a bad credit risk. **Responsibilities:** 

- Designed Logistic Regression Model.
- Created Confusion Matrix to have a proper visualization in accuracy of prediction.
- Attended AUC-score of 0.77 with Logistic Regression model and predict the customer is good credit risk or bad credit risk.

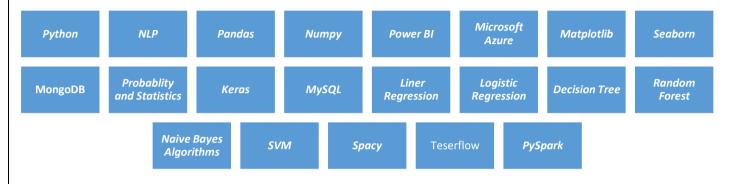
## **Title: Air Quality Index Prediction**

## Responsibilities:

**Description:** To develop an end to end model that can be used to predict the index quality of each day of a year.

- Developed a web scrapping tool that collects air quality data from various resources via URL link.
- Pre-processed the data to remove null, missing and inappropriate values.
- Designed Linear Regression model, Ridge regression model, Lasso Regression model and Random Forest Repressor.
- Used GridSearchCV to get the best combination of hyper-parameters for every model used
- Executed various performance measures including R-square & mean-squared error to select the best model.
- Attended accuracy of 79% with Random Forest Regressor model and used this model to predict the air quality.

## TECHNICAL SKILLS



#### **PERSONAL DETAILS**

**Date of Birth:** 15<sup>th</sup> December 1994 **Languages Known:** English and Hindi

Kaggle Id:https://www.kaggle.com/singhbaijuLinked Idbaiju-kumar-singh-881803115/

Address: Mjr Pearl Flat no 301, I Tower opposite Whitefield railway station Bangalore