# Ashish Kumar

📞 (+91) 8861302568

✉ ashishkr.568@gmail.com

📍 Pune (India)

in https://www.linkedin.com/in/ashish568/

○ https://github.com/ashishkr568

🌐 https://ashishkr568.github.io/

## ABOUT ME

Data Scientist with experience in providing insightful data-driven solutions to customers in multiple domains such as energy, media, networking, healthcare and construction. These end to end solutions includes data mining, ETL tasks, data exploration/visualization, predictive modelling (using statistical methods and machine learning algorithms), model deployment and its maintenance/retraining. To know more about me, please visit my website.

## EMPLOYMENT

**Persistent Systems [Aug 2018 – Apr 2022]**

Lead Data Scientist, Pune India                                                              Aug 2018 – Apr 2022

**Tata Consultancy Services [Jan 2012 – Jul 2018]**

Data Scientist, Pune India                                                                   Oct 2014 – Jul 2018
Data Analyst, Bangalore India                                                            Jan 2012 – Sep 2014

## KEY PROJECTS

### World's First Combustion Tuning Machine Learning Model of 800 MW Coal Fired Utility Boiler

Develop an ML-based Digital Twin of 800MW coal-fired utility boiler to reduce time and cost for Combustion Tuning, prescribe input settings to meet emission norms for flue gases and get real-time prediction for key parameters.

- Performed ETL task on data from ~1500 sensors.
- Performed data cleaning activities such as handling missing values, outlier analysis and feature engineering.
- Performed Exploratory Data Analysis to generate insights from data and validated it with domain experts.
- Created data-driven models using R and Python for ~30 KPIs with real-time prediction.
- Optimized settings for efficiency during combustion tuning.
- Handled a team of UI developers, integrated the complete system and deployed the solution in real environment.

Reduced Combustion tuning time and cost by 80% and deployed the solution at 2 of the power plant in Japan and Taiwan.

### Email Relevancy Application

Develop a python-based application to filter and extract data from online content (60 shared mailboxes) and deploy the solution on Microsoft Azure.

- Extracted emails from 60 shared mailboxes using exchangelib and applied business rules to check their relevancy.
- Analyzed email body and subject to reduce Azure translation cost by 98% with an accuracy of 99.7%.
- Created Power BI dashboard to monitor application performance.
- Developed multiple logic apps to automate manual tasks.

This application reduced manual efforts of classifying emails by 85%.

### Predict promotional discount applicability on a deal

Predict if a promotional discount will be added to the deal for a networking company.

- Perform ETL task on customer sales data for 1 year.
- Identified promotional discount trends across regions.
- Predicted promotional discount percentage for each deal.

Predicted promotional discount percentage with an accuracy of 97%.

### Inventory Management and Sales Forecast for Alcoholic Beverages

Create a case study for Inventory management and sales forecasting of alcoholic beverages across a US state.

- Used a public dataset and classified beverages into 4 major categories
- Identify inventory trends of each category to get an idea about inventory management of a new store
- Forecasted sales of these beverages for a dealer.

Achieved an accuracy of 91%.

### Duplicate Report Identification

Identify duplicate and similar reports to reduce migration efforts for a payroll service company.

- Extracted reports metadata from XML format.

- Used cosine similarity and a custom algorithm to identify duplicate and similar reports.

Reduced report migration efforts by 20%.

## Exploratory Data Analysis (EDA) to identify the root cause of dialysis failure

Identification of root cause for dialysis failure for a dialysis instrument manufacturing company.
- ETL operations on data of 100 dialysis sessions from 10 different machines.
- Visualize trends across failed dialysis sessions and identified the root cause.

Received appreciation for identification of root cause and visualizations used to explain the findings.

## Credit Insurance Bot

Develop a chatbot to answer queries related to credit insurance to reduce the load on customer care.
- Created a chatbot using Microsoft Cognitive Services such as LUIS and QnA Maker.
- Deployed the solution using Microsoft Bot Framework and Bot Composer.
- Provided training to the customer for future enhancement of corpus and the addition of new intent.

This bot is successfully deployed on the customer's website.

## HONOURS AND AWARDS

| | |
|---|---|
| **Bravo - Individual Awards [Oct 2021]** | Excellent work ethics, delivery beyond expectations and availability round the clock. |
| **Bravo - Individual Awards [Jun 2021]** | Development and deployment of Chatbot for a customer in the Credit Insurance domain and managing the customer independently. |
| **High Five - Individual Awards [Aug 2020]** | Excellent work and presentation on Chatbot. |
| **Best Team Award [Jul 2018]** | Successful delivery of the project and achieving a customer satisfaction index of 100% |
| **Certificate of Commendation [Nov 2017]** | Efforts towards successfully developing and implementing the World's First Combustion Tuning Machine Learning Model of Mitsubishi Hitachi Power System's (MHPS) Ultra Super Critical Coal Fired Boiler. |

## DIGITAL SKILLS

| | |
|---|---|
| Cloud Computing | Microsoft Azure | Amazon Web Services (AWS) (Basic) |
| Programming Language | Python | R | VBA for Excel |
| Database | Microsoft SQL Server | PostgreSQL |
| Statistics | Predictive Analytics | Inferential Statistics | Descriptive Analytics |
| Modelling Technique | Statistical Modelling | Machine Learning | Deep Learning |
| Machine Learning | Supervised Learning | Unsupervised Learning | Semi-Supervised Learning |
| Algorithms | Decision Trees | Random Forest | XgBoost | Linear Regression | Multivariate Adaptive Regression Splines (MARS) | Partial Least Squares (PLS) | K-Nearest Neighbors (KNN) | Logistic Regression | Support Vector Machines (SVM) | SARIMAX | Artificial Neural Network (ANN) (Basic) | CNN |
| Unsupervised Learning | K-Means Clustering | Hierarchical Clustering | DBSCAN | Principal Component Analysis |
| Reporting Tools | Power BI |
| NLP | ChatBot | Rasa | Text Classification | Sentiment Analysis |
| Data Science Libraries | Pandas | Numpy | Matplotlib | Seaborn | Scikit-learn | NLTK | Tensorflow | Keras | Vader | Tkinter (GUI) | pyPDF2 | Dask |
| Web Development | HTML | CSS |

## EDUCATION

| | | |
|---|---|---|
| **University:** | Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India | Jul 2007 - Jul 2011 |
| **Degree:** | Bachelor of Technology (Electrical Engineering) | |

## LANGUAGES

English |  Hindi