

ASHUTOSH KUMAR

+91-8792482248 • Bangalore

g.oneashu@gmail.com • <https://github.com/akum001>

SUMMARY

- 2+ years of experience as a **System development associate/Machine learning** with a proven track record of develop and implement large scale algorithms that have significantly impacted quality of the services.
- 2 years of experience in applying wide range of functions such as organizing data, understanding data, data pre-processing, data visualization, predictive modelling, content discovery and NLP.
- Hands-on experience in applying several ML/Statistical algorithms such as **Linear Regression, Logistic Regression, SVM, Random Forest, Clustering, Neural Networks, Gradient Boosted Trees, Stochastic Gradient Descent, Xgboost.**
- Hands-on experience in using **Regular expression.**
- Performing Data Analysis using **SQL** and **Excel.**
- Hands-on experience in applying **NLP** using **NLTK** and **Scikit-learn library.**
- Hands-on experience in applying **Ensemble Techniques** and **K-Fold CV** to achieve good results.
- Hands-on skills at programming in **Python** and **Java.**
- Excellent business communication skills.

TECHNICAL SKILLS

Programming Languages	:	Python 3.5 , Java 1.7
Databases	:	SQL, Sqlite
Operating Systems	:	Windows 95, 98, XP, Vista, Windows 7, Ubuntu
Other Technologies	:	HTML, CSS, Excel
Tools	:	Git, Perforce 1.4
Analysis library	:	Pandas, Scipy, Numpy
Cloud	:	Amazon EC2
Visualization library	:	Matplotlib, Seaborn
ML library	:	Scikit learn, Tensorflow
Machine Learning	:	Linear regression, Logistic regression, KNN, Decision trees , Naive Bayes, SVM, TF-IDF, K-means clustering, SVD/PCA, Random forest Neural networks
Statistics	:	Hypothesis testing, Bayesian inference

WORK EXPERIENCE

System Development Associate (Associate Analyst)
Xerox Business Services Ltd – ITPL, Bangalore.

April 2016 to Present

Project #1

Title : **Multiclass Classifier**
Client : Windstream and GTT

Technology Used: Python, Numpy, Pandas, Matplotlib, Scikit, NLTK, Jupyter Notebook

ETL framework is used for development and standardization of parsers. It automates the invoices processing system. The system is used to convert the e-bills to BDT-SECAB or NEXTGEN format by using ETL framework.

It generates outputs in the form of BDT or CSV and follows SECAB or NEXTGEN billing standard as developed by Alliance for telecommunications Industry (ATIS) to enable billing records to be exchanged between small service providers also called vendors.

Build a category classifier that had four classes named MRC, OCC, SURG, and TAX, to predict which charge belongs to which class based on charges description. It improved the quality of output files and reduced the volume of tickets.

Roles and Responsibilities:

- Data preparation and exploratory data analysis.
- Noise removal including language stopwords, punctuations and industry specific words.
- Tokenizing corpus and creation of Bag of words.
- Conversion of Bag of words into Vector form.
- I used TF-IDF to convert corpus vector to term frequency and inverse document frequency.
- I used Naive Bayes classifier to train and test the model.

Project #2

Title: Customer churns prevention

Client : Windstream

Technology Used: Python, Numpy, Pandas, Matplotlib, Seaborn, Scikit, Jupyter Notebook

Customer churn refers to when a customer ceases his or her relationship with a company. I came up with a model that helped my client to predict if any customer is going to churn so that they can prevent this before it happen

Roles and Responsibilities:

- Exploratory data analysis, data cleaning, feature importance analysis, fill in missing values.
- I used cross validation where number of folds was 10 to reduce the chances of over fitting.
- Trained the model with different test sets and used average testing accuracy as the estimate of out of sample accuracy.
- I used KNN, SVM and Random Forest to train my model and selected the classifier which had highest accuracy and F1 score.

EDUCATION & CERTIFICATION

Bachelor of Computer Application Chaudhary Charan Singh University, Meerut	2012 – 2015
---	-------------

Stanford online certification in Machine learning	2017
Udemy certified in Python for data science and Machine learning	2017

PERSONAL

Date of Birth	:	12/11/1992
Citizenship Natural	:	India
Languages Interests	:	English, Hindi

DECLARATION

I hereby declare that the information furnished above is true to the best of my knowledge.

Thanking You

- Ashutosh Kumar

Place: Bangalore