KOUSHIK KHAN

IBM India Private Limited, Embassy Golf Links, Bangalore - 560071, India

Cell: (+91) 7501-923-251 • **Email:** koushik.khan@in.ibm.com • **LinkedIn:** https://www.linkedin.com/in/koushikkhan/

ABOUT

Koushik is currently working as a Data Scientist at IBM. He has more than three years of experience in industry in the field of Predictive Modeling, Machine Learning and Information Retrieval. He completed his graduation and post-graduation in Mathematical and Computational Statistics from Visva-Bharati University, Santiniketan, India.

He was selected as DST Research Fellow by the Department of Science and Technology, Government of India for outstanding academic achievement in his Masters program. Koushik takes special interest in advanced analytics, machine learning, image and language processing.

EDUCATION

Visva Bharati University, Santiniketan *June, 2013 - June, 2015*
Master in Statistics
Department of Statistics
Thesis Title: Joint Modeling of Longitudinal and Time to Event Data

Visva Bharati University, Santiniketan *July, 2010 - May, 2013*
Bachelor in Statistics
Department of Statistics
Thesis Title: Method of Estimation, A Generalized Approach

KEY SKILLS

- Applied Statistics and Predictive Analytics with R and Python
- Unstructured Data Analysis & Natural Language Processing
- Information Retrieval
- Machine Learning

PROFESSIONAL EXPERIENCES

Marketing Mix Modeling (MMM), Nestle

Description: (i) Developing marketing mix model for Indonesia and Pakistan markets to understand their sales pattern by identifying the contributions of several key marketing factors like TV GRP, FACEBOOK/YouTube Impressions on Nestle products etc. (ii) Developing Optimization module to find the optimal contributions under critical marketing constraints on varoius factors. (iii) Developing an interactive Simulator to visualize what-if scenarios from the model.

Contribution: (i) Contributed in designing modules for data processing pipeline using R and Azure SQL Server. (ii) Developed several R packages to cater the need for automating the computational processes before developing the MMM models. (iii) Contributed in designing the overall workflow.

Application Monitoring Automation - NLP with IBM Watson and Python, IBM

Description: (i) Creating a custom entity recognition model using IBM Watson to identify **problem**, **ask**, **user** and **object** type entities from customer complaints data. (ii) Creating an unsupervised ML model to identify possible root causes from available customer complaints. (iii) Creating a complete workflow by extracting investigations, observations and resolutions, after having identified the root causes in point (i).

Contribution: (i) Contributed in the data processing pipeline using Python before sending it to IBM Watson for training, it includes dictionary creation by identifying key-words, creating rules to identify users by their professional email addresses etc. (ii) Used Latent Dirichlet Allocation based topic model to identify possible root causes. Also contributed in deployment by creating REST services in Python. (iii) Contributed in designing source codes and implementation for identifying investigations, observations and resolutions oriented textual chunks by generating grammar based patterns. Also contributed in code integration and deployments.

Contextual Search - Information Retrieval, Barclays

Description: Barclays online help system used to run on keyword based search algorithm which made the search results less accurate. The objective was to apply information retrieval techniques to make the search engine contextually more intelligent. We designed and developed such a model based on matrix factorization techniques to overcome the problem.

Contribution: Designed a customized SVD based information retrieval algorithm for Barclays help Q&A data set. Implemented the same in Python.

Cognitive Testing for IBM Ignite - NLP with Python, IBM

Description: (i) This project (Test automation program) was all about building a cognitive solution for software bug tracking and throwing recommendations in case of any issues. The specific role is to collect data from the client, build algorithms and implementations of those in a production environment.

(ii) Building defect classifiers (classification models) to classify a software bug into one of the various standard categories by analyzing problem descriptions, titles and providing end-to-end support.

(iii) Building application programming interfaces to provide ways to interact with the models.

(iv) Building some task automation tools in Python that can help the users to configure their own systems for using the Ignite System.

Contribution: (i) Contributed in developing defect classification models by analyzing complaint texts, their associated REST APIs' and in providing end-to-end support. (ii) Contributed in developing algorithm as well as cognitive tool for Ignite Defect Analytics that was supposed to show recommendations to the testers and testing leaders in the presence of software bugs. (iii) Contributed in integrating cognitive tools to the entire system by developing various task automation scripts in Python.

Computation of Data Compression Rate in Lossy Wireless Network Communication - BCausE R&D Lab

Description: Developing mathematical rules for computing the likelihood for several wireless network objects like multiple GSM SIM slots and wifi module for data flow. This includes collection of data on several wireless network related variables like throughput, packet loss etc., data normalization, application of classification algorithm like k-nearest neighbor.

Contribution: Contributed in developing methods for computing the likelihood for several wireless network objects like multiple GSM SIM slots and wifi module for data flow using classification model.

AWARDS & RECOGNITION

- Achieved **Scholarship for Higher Education (SHE)** in 2011, sponsored by Department of Science and Technology, Govt. of India.
- Achieved first position in B.Sc. Examination (Statistics), 2013.

- Achieved Second position in M.Sc. Examination (Statistics), 2015.
- Achieved **DST Fellowship** to pursue research for excellent performance in M.Sc., 2016

PERSONAL PUBLICATIONS

- A Not So Short Introduction to Object Oriented Programming using R
- Demonstration of Central Limit Theorem with Simulated Binomial Variates
- Effect of BoxCox Transformation on Non-Normal Data Sets
- Principal Component Analysis – An Introduction with R Implementation
- Basic Data Manipulation with dplyr

OTHER JOB RELATED ACTIVITIES

Interested in studying and building machine learning algorithms, coding and blogging related to data science.

Github Repositories:

- **https://github.com/koushikkhan/The-Gradient-Descent-Optimization**
- **https://github.ibm.com/koushik-khan/Guided_Topic_Model_1**
- **https://github.ibm.com/koushik-khan/ClassificationPipeline**