Nashit Babber
Co Founder and Data Scientist
Hezkrost Technologies(Company Closed)
Pune, India
Mobile : +91 - 8421000030
Email : nashit93@gmail.com

Profile :

Business minded , Coursera certified data scientist and machine learning engineer with demonstrated ability to mine hidden gems in the ocean of data.

Overview :

➡ 2.5 Years of extensive experience in Data Science and Big Data Technologies on both on-premise and cloud Amazon Web Services environment.
➡ 17 certifications from Coursera on Data Science , Machine Learning and Artificial Intelligence
➡ Experience in Machine Learning with emphasis on unsupervised learning, zero knowledge analysis , model-free and feature free classification , non-parametric, non-linear modelling , classifications and predictions
➡ Experience with designing high capacity trading strategies targeting the equity market.
➡ Experience in working with big data problems arising in Life Science Domain.

Key Skills :

Software Library	Tensorflow
Languages	Python , Shell Scripting , R , C++
Python Libraries	pandas , scikit-learn(sklearn) , MatplotLib , scipy , numpy , tflearn , nltk , requests , bs4 , selenium , lxml.
Hadoop EcoSystem	PySpark, MapReduce2, HDFS , Hive , Sqoop , Impala ,Cloudera(CDH)
DataBase(s)	MySQL, SQLIte, CouchDB
Cloud	Amazon AWS/EC2
Platforms	Linux , Mac OS , Windows

Work Experience :

Co-Founder and Data Scientist
Hezkrost Technologies(April 17 - till date)

Executive Summary :
Creating stock prediction system from scratch using machine learning and deep neural networks.

Roles And Responsibilities :
‣ Preparation , planning , setup and installation of cost effective ways to extract data from NSE website.
‣ Automated web scrapping to extract news about all 1857 stocks from google finance.
‣ Automated web scrapping for fetching daily historic data, bhav copies, daily high/low , Advances and Declines, Bulk Deals of stocks
‣ Automated scraping of realtime data from NSE websites
‣ Automated scraping of Income Statements, Balance Sheets, Cashflows from Yahoo Finance.
‣ Fresh loading and incremental loading of data into HDFS.
‣ Data cleansing and validation of daily updated data.
‣ Training , Applying various Classification algorithms and prediction on Balance Sheets, CashFlows and Income Statements to classify stock fundamentals as week or a strong.
‣ Applied Time Series analysis over the historic data of stocks
‣ Created hive scripts to monitor the daily advances and declines, Bulk deals , daily high/lows.
‣ Training and predicting stocks prices based on certain parameters of historic prices using regression algorithms
‣ Assisted in sentimental analysis of stocks news data to categorise the stock sentiment as positive, negative or neutral.
‣ Recently started applying LSTM Recurrent Neural Network to near-real time data being extracted through NSE website.

Hadoop Developer And Data Scientist
Saama Technologies(Dec 16 - Apr 17)

Executive Summary :
Migrating fresh and incremental data from Teradata to HDFS and creating scripts for Extraction , Transformation and Loading using pyspark.

Roles And Responsibilities :
‣ Collaborated with onsite team for planning and production process of data migration.
‣ Involved in creating scripts for sqooping historic and daily shipment data from teradata using JDBC drivers.
‣ Creating shell script for validating daily sqooped records and storing the counts in a file
‣ Created a shell script to send email with the summary of daily load , script failure messages and logs generated on failure.

- Created validation script in python which took sample of the incremental load and validated its format and data types and generated logs accordingly.
- Created transformation scripts to join data from different shipment companies (such as FedEx , UPS , BlueDart etc) with the Tailored Brand's database and find out the shipment details , delays in shipments and other meaningful insights and display it onto tableau dashboard.
- Calculating the average shipment delay time by different shipment companies on seasonal basis and forecasting the number of days Tailored Brand should proceed with shipment for timely delivery in various regions.

Hadoop Developer
Saama Technologies(Oct 2015 - Dec - 2016)

Executive Summary :
Assisted in developing 4 Real World Evidence apps used for analysing patient's data and creating meaningful insights out of it. These apps included Cohort Builder , Treatment Pathways , Population Characteristics and Market Focus.

Roles and Responsibilities :
- Designing and planning of the backend structure and data flow for all the 4 apps.
- Assisted in creating impala and hive scripts for backend data loading of all the four apps.
- Debugging the HQL query issues and improving the query performance during the design process.
- Created python scripts for automatically executing queries for Extraction ,Loading and Validation of data in backend and emailing the summarised results to the manager.
- Created scripts in python for automatically converting any HQL scripts into XML files which were later required in Saama's Fluid Analytics Engine workflows.
- Created python scripts for complex data extraction with validating(and making probable changes where ever required) in each and every row loaded in hive tables.
- Scraping diagnostics data from the web and saving it into csv files using selenium drivers.
- Cooking random data in python for development stage of the project.
- Worked with different compression(Such as parquet, snappy) methods for optimising the storage.
- Created an algorithm to calculate distance between two cities in US where clinical trails can be conducted and grouped them according to their distances from each each other.

Certifications :

Name Of The Certificate	Specialisation	Certifying Authority	Certificate Link(With Grades)
Data Science Specialization	**Data Science**	**Coursera**	**Click Here!**
The Data Scientist's Toolbox	Data Science	Coursera	Click Here!
R Programming	Data Science	Coursera	Click Here!
Getting and Cleaning Data	Data Science	Coursera	Click Here!
Exploratory Data Analysis	Data Science	Coursera	Click Here!
Reproducible Research	Data Science	Coursera	Click Here!
Statistical Inference	Data Science	Coursera	Click Here!
Regression Models	Data Science	Coursera	Click Here!
Practical Machine Learning	Data Science	Coursera	Click Here!
Developing Data Products	Data Science	Coursera	Click Here!
Data Science Capstone	Data Science	Coursera	Click Here!
Machine Learning(By Stanford and Andrew NG)	**Machine Learning (11 Weeks Course)**	**Coursera & Stanford**	**Click Here!**
Deep Learning Specialization	**Deep Learning**	**Coursera & Deeplearning.ai**	**1 Pending**
Structuring Machine Learning Projects	Deep Learning	Coursera & Deeplearning.ai	Click Here!
Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization	Deep Learning	Coursera & Deeplearning.ai	Click Here!
Neural Networks and Deep Learning	Deep Learning	Coursera & Deeplearning.ai	Click Here!
Convolutional Neural Networks	Deep Learning	Coursera & Deeplearning.ai	Click Here!
Sequence Models	Deep Learning	Coursera & Deeplearning.ai	Pending
Introduction To Deep Learning	**Advanced Machine Learning**	**Coursera & National Research University Higher School of Economics**	**Pending**

Academic Qualifications :

Examination	College	University/Board	Percentage	Year Of Passing
BE(Mechanical)	PES Modern College Of Engineering, Pune	Pune University	61%	2015

Extra Curriculum :

‣ Technical Blogger. Check out my blogs here!
‣ Badminton & Table Tennis player.
‣ Assassin Creed addict!
‣ Creating Home Automation project on Raspberry Pi projects.A trailer here!
‣ And of course I love to travel.

Volunteer Works :

I am co-founder of freejee.com. Although at a early stage, We are dedicated to provide free education across the globe for all level of students(**Absolutely Free! No Copy Rights or Sign ups!**).

You can check out my current work Here!

Links :

Github : https://github.com/nashit93
Stackoverflow : https://stackoverflow.com/users/6323230/nashit
LinkedIN : https://www.linkedin.com/in/nashit-babber/