NILANJAN CHATTERJEE

Germantown, Maryland

Data Scientist @ Hughes |H1B| DS | ML| AI | Deep Learning| Python| R | SQL| Hadoop | Spark| Kafka| Recommendation | CD

arkasuperstar@gmail.com - +1 704-345-9929

Summary

Result oriented Data Scientist with 5+ years of experience in Data Science/ Machine Learning/ Data Engineering with Retail,BFSI, Insurance and Telecom domain skills.

• Extensive experience in Python and R for building Machine Learning pipelines by processing large volumes of data from varied sources and processing them as per business requirements using R(kernlab,doMc,shiny,ggplot2), Python(scikit-learn,numpy,scipy,pandas, matplotlib, NLTK,AllenNLP ,Spacy,Gensim), Spark, Hadoop,Solr, SPSS, SAS E-Miner and generating visualizations using Tableau/Qlikview/D3.

• Proficient in statistical tests, Hypothesis testing, Random effects,t/z-test, ANOVA, ANCOVA, Multivariate Analysis, Market Basket Analysis, ARIMA models, Text Analytics(NLP) | GPU Prog(CUDA) | Supervised(SVM, Neural n/w) and Unsupervised(PCA, NNMF, SVD,Clustering) algorithms.

• Software Developer, with extensive experience in design, development and testing in SDLC with experience in DevOps & Agile using Deployment, Build and Release with CI using Chef, Puppet, Ansible, Docker, Kubernetes, Maven, Git, Jenkins,.

Technical Skills :

Languages: C, C++, R (dplyr,e1071,caret,ggplot,nnet), Java, Python (numpy, pandas, scipy, sklearn, matplotlib, Seaborn, Spacy, Gensim, Spacy,H20,Vowpal Wabbit, PyTorch, Keras, TensorFlow, Caffe), SAS, Orange, WEKA,, Rust, Octave, LISP.

Databases: Microsoft SQL, MYSQL, Oracle, Mongo DB, DynamoDB, HBase

Tools: Tableau, QlikView, Looker,Tableau, SAS 9.4, OpenCV, Hadoop, HBase, Oozie, Pig, Kafka,, CPLEX, Gurobi.

Data Science: Classification, Regression, Feature Engineering, Clustering, Data Mining, Predictive Modelling, Deep Learning,

Statistics: Time Series, Hypothesis testing, Information Retrieval, PCA, Dimension reduction,Recommender Systems,

Cloud: AWS, GCP, Azure,

Reachable @ +1-704-345-9929/nilanjan.9325@gmail.com

Experience

Data Mining Scientist at Hughes
June 2018 - Present (1 year 4 months)

- Dashboard and reporting stack development using Tableau, SAP Web Intelligence, Qualtrics, Oracle, SQL Server and Hive to cater multiple reporting requirements for global markets.
- Working for anomaly and outlier detection for network usage data and integrating it to Tableau reports for QA team and improving Cx experience by enhanced Sentiment Detection, Sarcasm Detection, Mood/ Topic/ Priority detector using Clarabridge Analytics, Python, Azure ML Studio, SAP Infinite Insights and RapidMiner.
- Predictive Analytics on WiFi data implemented usng Hadoop 2, Spark 2.3, Kafka, Hive. R, Python code repository in GitLab enterprise and reports using Tableau and SAP Business Objects. Also wrote custom Python scripts for data validation and PII masking.

Data Scientist at Advance Auto Parts
February 2018 - June 2018 (5 months)

- Wrote sqoop jobs for importing data from Oracle tables into HDFS and wrote Hive queries and User Defined Functions for managing/wrangling data using HBase(Tez) and Impala. Used Docker containers to synchronize code changes and stored the data in S3 buckets. Data from S3 was used for analytical processing using AWS Kinesis, Athena, Lambda and SAS VA, SAS DI.
- Worked on implementing ARIMA with exponential smoothing and in addition to Kalman filter for better prediction accuracy. Also working on k-NN, mixed model ANOVA, Poisson regression in SPSS modeler, Python for product assortment.
- Wrote SAS macros and SAS SQL jobs to perform data aggregation and feature scaling.
- Set up Airflow DAG's and fork scripts for local and Mesos executor to schedule python jobs to run weekly and update S3 and AWS Batch. Manually scripted DAG of DAG's for process automation and faster data updation by parallelizing data update process by integrating AWS plugin in Python.
- Created workflow automation using Docker swarm on Elastic Container Service and AWS Batch. Integrating Elastic search, Logstash into the Order Management system and scheduling Oozie jobs for database update.
- Worked for minimizing anomaly detection for retail product assortment. Also working on NLP models for Topic modelling by POS tagging and feature extraction from review data for building recommendation engine by collaborative filtering and deployment of the models using AWS Glue and Sagemaker.
- Built visualization dashboard integrating S3, Tableau with Hive connector and SAS Viya.

Data Science Intern at IBM
May 2017 - January 2018 (9 months)

- Developed solutions for BlueMix team by feature engineering of usage pattern data/ customer churn & developing algorithms like PCA, SVD, Market Basket Analysis, kNN, GBM, Hierarchical Clustering, Logistic Regression with Bayesian Belief, MCMC.
- Worked with HR Analytics team implemented NLP, IR, ranking algorithms to predict employee attrition rate and identify them.
- Have designed model deployment and optimization strategy using CPLEX, Gurobi, LAPack for Spark MLib platform.
- Have created visualizations using Python (Matplotlib, Seaborn), R(ggplot2) and effective roll out and pilot testing strategy with Sales team based on model AIC performance and iteration strategies for A/B testing.
- Worked on Building hyper ledger models for Predictive Insights for Block Chain team on Blue Mix cloud SaaS offering for smooth process transitioning into secure distributed ledger distribution network.

Senior Data Engineer at Hewlett Packard Enterprise
May 2016 - August 2016 (4 months)

- Led a team of three towards implementing statistical tests, linear contrasts, GBM, Random forests, Monte Carlo for HP scanner devices.
- Have worked on Time series modeling for sales/demand forecasting implementing ARIMA, GARCH models with seasonal and exponential smoothing.
- Have wrangled huge terabytes of data using R(dplyr,tidy), OpenRefine, Trifacta to enable ML algorithm development.
- Developed predictive models using R, Python, Scala and presented on scrum meetings for consumer analytics. Had setup 10 node SAS SPD/Spark cluster for stream analytics of pilot phase using Spark, Kafka, Flume Mahout, Impala, HBase, Hive Oozie.
- Worked on Azure and AWS for cloud integration using S3, EBS, EMR, Dynamo, RedShift, Lambda, Lex following CRISP-DM.

Data Engineer at Cognizant
October 2014 - May 2016 (1 year 8 months)

- Have written efficient advanced SQL queries and cubes using SSIS, SSAS, SSRS, PISQL & T-SQL, with Microstrategy MDX query.
- Qualitative data ,sentiment analysis and Topic modelling & build NLP models and visualizations using R(topicmodels, lda, tm, e1071, tidyverse, plyr) .
- Have migrated DW's and data lakes to 20-25 node Hadoop cluster with Pig, Hive, HBase, Sqoop, Impala, Oozie, Flume, Ambari, Mahout. Have set file formats used Avro,ORC, Parquet and also setup authentication using Ranger ,integrating Kerberos with LDAP.
- Have worked extensively on SQL development and created triggers, functions, cursors, packages and defined execution plans

• Worked on DW integration and management using Erwin Data Modeler, MSSQL, Power BI, Pentaho, Cognos, UNIX. Have experience managing lifecycle of application deployment by technologies like SQL Server, SSIS,, SSMS, Toad, Unix, Putty

Programmer Analyst at Cognizant
July 2013 - September 2014 (1 year 3 months)

- Multi-channel attribution analysis as per given dataset from clients and analysing using available toolkits like Looker and Google Analytics integration with QVD.
- Have worked on Docker, Vagrant, Nginx, Jenkins, Git , UrbanCode ,Bugzilla, Ansible, Chef, JIRA, Microsoft TFS for project deliverables.
- Created and deployed MSI and virtual applications using SCCM 2012 , Landesk and Altiris . Customized Mobile Device Management and BYOD implementation using Air Watch, Intune, and Mobile Iron.
- Quality Analyst and Deployment Lead for Application Packaging and Deployment , OSD Imaging, Patch management using SCCM 2007/2012/Altris/LanDesk/Marimba with hands-on scripting using VBScript,Powershell, Bash, Python, Perl.

Education

University of North Carolina at Charlotte
Master's Degree, Computer Science, 2016 - 2017
Activities and Societies: Triveni Intramural Sports
NARULA INSTITUTE OF TECHNOLOGY
Bachelor's Degree, Electrical, Electronics and Communications Engineering, 2009 - 2013
Activities and Societies: IEEE
VIVEKANANDA MISSION SCHOOL
ISC, Physical Sciences, 2007 - 2009
Activities and Societies: SPICMACAY

NILANJAN CHATTERJEE

Germantown, Maryland

Data Scientist @ Hughes |H1B| DS | ML| AI | Deep Learning| Python| R | SQL| Hadoop | Spark| Kafka| Recommendation | CD

arkasuperstar@gmail.com - +1 704-345-9929

Linked in. Recruiter

1 person has recommended NILANJAN

"Nilanjan is calm and composed. He is extremely detail oriented and analytical. He is meticulous learner. This makes a sought-after resource for any complex piece of project.He would definitely make a great addition to any team engaged in complex software development."

—Shubhodaye H G, Test Engineer, Infogain, managed NILANJAN indirectly at Hewlett Packard Enterprise