# 

# Arun Kumar Bonam

# Mail Id: arunkumarbonam01@gmail.com Mobile: +91-9700433068

# 

# SUMMARY OF EXPERIENCE:

# Arun has 3.6 years of professional experience (offshore and onsite) in the IT industry.

# Good knowledge and hands-on experience in developing different Machine Learning algorithms using python libraries like pandas, numpy, matplotlib, scikit etc.

# Experience in implementing different machine learning algorithms like Logistic Regression, Classification, Random Forest, Decision Tree Algorithms using Python.

# Experience in building Artificial Neural Networks using Deeplearning libraries like Tensor flow and keras libraries in python.

# Experience in building processes to source and process data from variety of sources data to and from Hadoop systems using Hadoop stack and Spark with Scala.

# Experience in Building distributed real-time stream processing and batch processing using In-memory Data processing technologies like Apache Apex (distributed in memory processing with Java).

# Experience in Developing Applications which uses NoSQL databases like Hbase.

# Experience in optimizing performance of applications by understanding the memory mapping of cluster.

# Experience in development of Big Data ETL Applications using Hadoop Stack-Map Reduce, Hive, Sqoop, Flume, Apache Apex.

# Experience in designing scalable architectures for the big data solutions as per the requirement of client.

# Experience in development of java applications using Core Java, Servlets, and JSP and Tomcat Web server.

# Experience in Data Spooling from different source logs using Hadoop fault tolerant technologies like Flume.

# Experience in Migrating applications from traditional DataWare house to Hadoop platform.

# PROFESSIONAL EXPERIENCE:

* Working as Data Engineer with Groupon, Bengaluru India from August 2017 till date.
* Worked as Senior Systems Engineer with Infosys Limited from Dec2014 to August 2017.

# EDUCATIONAL PROFILE:

* **B. Tech** in Electronics and Communication Engineering from JNTU Kakinada, India.

# TECHNICAL EXPERTISE:

# 

# Data Science: Regression, Classification algorithms with Python pandas, scikit, numpy,

# matplotlib, Natural Language Processing with nltk, DeepLearning with TensorFlow and keras libraries.

# Big Data: Spark with Scala, MapReduce, Hive, Sqoop, Flume, Hbase, Apache Apex

# Hadoop Distributions: Hortonworks, IBM Big insights, Cloudera

# Java Technologies: Core Java, Servlets, JSP.

# Tools: Jupyter Notebook, Intellij, GitHub, Jira

# Database: Oracle, Netezza, Teradata

# Project # 1: GrouponIntelligence(GI)

# Client: GroupOn

# Role: Data Engineer

# Technologies:

# Python-Pandas, Numpy, Matplotlib, Scikit etc. Deeplearning libraries – TensorFlow, keras

# Hadoop Technologies-Spark, Hive

# Data warehouse: Teradata

# Description:

# Basically, this is not a single project. Our day to day activities will change

# According to the requirements from different teams.

# Activities includes solving various problems to extract meaningful data

# and generate actionable insights from wide variety of DataSources.

# We build different machine learning models based on the dataset and the requirement ranging from Decision Trees, Naive Bayes, KNN, SVM.

# We build DeepLearning models as well using Tensorflow and Keras to identify

# the deals which are not selling as expected and which are likely to gain more popularity.

# We solve problems includes various modules like anomaly detection in deals,

# Potential regions to expand the market, classifying the deals based on the region

# Etc.

# We use Python’s pandas, numpy, scikit, tensorflow, keras etc. and plotting various representable graphs using matplotlib.

# We also provide meaningful data to the downstream developers after cleansing and filtering the incoming data based on the decisions made by machine learning algorithms in identifying the required variables.

# Roles and Responsibilities:

# Typical day to day activities includes data loading from different sources which

# we will be using to create algorithms.

# Data cleansing and Data pre-processing activities like handling missing data, Feature Scaling, OnehotEncoding etc. using python’s libraries sklearn, numpy

# Identifying the best possible algorithms for the dataset and type of problem.

# Creating DeepLearning models like ArtifcialNeuralNetworks using tensorflow, keras

# Creating normal Classification, Regression models and identifying the performance of each model on the particular dataset.

# Creating visualizations for the models which we developed using python’s matplotlib.

# Adhoc activities like Loading the cleansed and filtered data into Hadoop and Teradata systems.

# Creating Hadoop jobs to load meaningful data into downstream Hadoop systems

# Using Spark.

# Project # 2: Mobile Transformation Programme (MTP):

# Client: Virgin Media

# Role: Senior Developer (Onsite)

# Technologies: Hadoop stack (Map Reduce, Sqoop, hive, flume), Hbase, Unix, Netezza, Control-M

# Description:

# This project includes Providing the solution to client to use their own newly built Next Generation Business System (NGBS) to maintain their customer information, and other

# stakes like billing, marketing, ledgers etc. The primary purpose of this project is to help the client to move their current systems which are running on top abinitio and Netezza to Hadoop systems which will further used for number of requirements like reporting, Marketing analysis and providing new offers to the customers and Simswap Campaigning, generating billing reports, Antifraud Systems (Internet Security) etc. Various technologies from Oracle, Netezza to advanced technologies like Hadoop stack (Hive, sqoop, flume, Hbase) are used.

# Key Role and Responsibilities:

# Designing and developing the Hadoop solutions to handle huge volume of data flowing into system from different sources which will then flow into data warehouse system for creating Reports and Data Analysis etc. Using Data Spooling and Data Lineage and reconciliation solutions available in Hadoop (flume, hive, hbase).

# Designing and developing the application on top of hbase which will be used by Client’s Internet Security team to detect Spammers on client’s network.

# Designing architecture and developing solution to migrate dataware house (Netezza) to Hadoop eco system to store and query for the data and then creating reports

# Closely working with client side Architects in building architecture for different

# Systems as part of migration of old system into Hadoop platform.

# Developing POCs to analyze the working capability of designed system before starting actual development and giving demo to the client.

# Project # 3: Supplier Management System (SMS):

**Client :** DirectV

**Technologies** **:** Hadoop- Hive, Hdfs, sqoop, Apache Apex (distributed processing with java), Core java

**Description :**

This program includes designing and developing payment processing system and Customer behavior analysis system on Hadoop platform. This includes designing and developing fault tolerant system which processes continuous and batch streaming of data in distributed environment using In-Memory parallel processing technique called Apache Apex which runs on Hadoop and Java platform.

# This System is to analyze the customers data available in different systems and in different levels and process it based on the rules defined and and push the processed data into MediaMastero where we can analyze the processed data for the payments to the client’s subscribers. The system has three phases:

# 1) Scheduling Sqoop jobs to pull the data available in different systems into hdfs.

# 2) Processing the data available in hdfs using Data Torrent RTS (Apache Apex).

# 3) Generating the results in Hdfs and sqoop the data to oracle by workflow scheduling.

# Key Role and Responsibilities:

# Lead one of the module in project and closely work with the rest of the module teams in the project.

# Responsible for developing the processing logic using Data Torrent Rts/apache Apex which runs on top of Hadoop and Java.

# Development of workflow which schedules sqoop job to bring the data from different sources into HDFS.

# Moving the data from HDFS to oracle using sqoop jobs after applying the business logic.

# Understanding the clusters behavior to optimize the performance of the application by using Jstacking techniques.

# Writing Junit test cases

# Fixing the bugs in unit testing/SQA and UAT stages.

# Knowledge transfer activities to the team.