Dhanendra Singh
Contact No: 860-989-3522 **Current Location:** Hartford, CT
Email id: dhanendrasingh@gmail.com

INTRODUCTION AND OBJECTIVE:

- **13+** years of extensive IT experience of which **9+** years in United States in the field of **Analytics, Design, Development** and **Management of Healthcare insurance applications** with expertise in **Claim Analytics and Adjudication.**
- Currently pursuing **Master's in Computer Science – Data Science** from University of Illinois at Urbana Champaign. Expected completion date is May 2019. **Current GPA 3.90**.
- **4.5+** years' experience working as a **Data Science/Data Analytics consultant** for Infosys's Data and Analytics unit (DNA Unit).
- Experience in **Exploratory Data Analysis, Data Cleaning, Data transformation, Data Integration** and **Data Visualization**.
- Skilled in **R**, **Python**, **SQL**, **Tableau**, **Machine learning techniques**, **Statistics**, **AI/ML platforms** and technologies.
- Knowledge on **NoSQL** Data extraction tools including **Hbase**, **Cassandra** and **MongoDB**.
- Experience in using various **packages** in R and Python.
- Experience in writing **Sub Queries, Stored Procedures, Triggers, Cursors using SQL**.
- Extensive experience in writing complex SQL statements using **unions, multiple joins, triggers and sub queries.**
- Knowledge on Hadoop Architecture and ecosystems such as **HDFS**, **Data Node**, **YARN**, **MapReduce**, **HBase**, **Hive, Pig**.
- Involved in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modeling and data visualization with large data sets of structured and unstructured data, created ER diagrams and schema.
- Knowledge on deploying and managing infrastructures based on **Dockers**, **Containers**.
- Used **Tensorflow packages** to train machine learning models as part of the academic projects.
- Good knowledge on **Pivot Table** reports in Microsoft Excel.
- In-depth knowledge of **healthcare systems and healthcare business from Payer perspective**.
- Experience working in **Agile** as well as **Waterfall** methodology.
- An enthusiastic self-starter with an excellent analytical and problem-solving ability.

Passionate about the application of Data Science to solve real world problems. Excited to contribute to the world of Data science, Big data and Analytics.

TECHNICAL SUMMARY:

- **Machine learning** - Regression (Linear and Logistics), Clustering, Decision tree, Random forest, SVM, KNN, Naive Bayes, Time series modelling, NLP, PCA, Regularization, Feature selection, Hypothesis testing, Model evaluation, Model optimization, Hyperparameter tuning, Ensemble methods like Bagging, Boosting, Stacking, Box-Cox transformations.
- **Data cleaning –** Regular expression, Open refine, YesWorkflow

- **Data curation**
- **Data Visualization** - Tableau, D3j, JavaScript, CSS
- **Cloud computing Application, Distributed computing** - AWS (EC2)
- **Tools and Technologies –** C, R, Python, HTML, COBOL, CICS, DB2, Hadoop, TensorFlow library, Hive, Pig, Flume, Spark, Anaconda, Juptyer
- **Database systems** - SQL, Relational Algebra, Neo4j, Mongo DB, HBase
- **File format** - XML, JSON
- **Domain –** Health Insurance (Payer)
- **Business Skills**: Business Definition Requirements, Business Process Analysis, Use Case Modeling & Analysis, Story writing.
- **Business Tools**: Rational Suite (Requisite Pro, Clear Quest, Clear Case), MS Visio, MS Access, MS Office Suite, RRC, RTC, Tableau, R, Advanced Excel – Pivot table and VLOOKUP, DB2 Studio
- **Other tools:** Xpeditor, Endevor, SDSF, Zeke, FTP, Fileaid, Plainview, Quickbase, QNEXT

EDUCATION:

- **Master's in computer science (Data Science) –** University of Illinois at Urbana Champaign. Expected completion date – May 2019. **Current GPA - 3.90.**
- **Tackling the Challenges of Big Data –** MIT Professional education, Digital programs – August 2015
- **Bachelor of Mechanical Engineering** - Completion date: June 2005. First class with Distinction
- **Diploma in Mechanical Engineering** - Completion date: June 2002. First class with Distinction

PROFESSIONAL CERTFICATIONS:

- Certified Scrum Master (**CSM**)
- American Health Insurance Program (AHIP): Fundamentals of Healthcare - Part A.
- AHIP: Fundamentals of Healthcare - Part B.
- AHIP: Healthcare Management AHM250
- AHIP: Accountable Care Act

ACHIEVEMENTS/AWARDS:

- Awarded Bachelor's in Mechanical engineering with First class and distinction
- Received **STAR** and **INSTA** and Knowledge Management awards from Infosys Limited
- Awarded **$2500 for IT Analytics Hackathon Gold award** (Visualization Category – Using Tableau) by **Aetna Inc. in 2015**.
- Worked on Multiple Request for Proposals, Consulting offerings and Solution accelerators for Infosys sales team.

PROFESSIONAL EXPERIENCE:

Company: INFOSYS LIMITED, USA

Job Title: Senior Consultant **Location: Hartford, CT**

Role: Data Analyst/Data Scientist **Client: Aetna**

June 2014 – Till Date

Projects:

Predict Medical cost - Using new data source and AI capabilities to more accurately predict Medical cost with 10% reduction in margin variance.

Identify Fraud, waste and Abuse - Discover incremental leads using AI and ML and wide range of internal and external data. Early identification of emerging fraud schemes using Machine learning models. Created Interactive visualization tool for quick review of claim and Medical record to identify abnormalities. Created Anomaly detection model by clustering individual providers based on procedure code.

Identify Overpayment and Payment errors - Identify hidden patters driving overpayment and payment errors using Machine learning models - Supervised machine learning model build tools to enable better identification and fix root cause of overpayment.

Created a statistical model (Linear regression using R) using historical data provided to help IT stake holders predict the project labor cost, defect rate and project success. Used Tableau to provide the UI Visualization that can be utilized by the stakeholder to evaluate their projects.

Responsibilities:

Extracted information from various internal sources, also procured data from external entity (This was done at a company level)
Worked with the Data Engineering team to ingest into Big Data platform from disparate data sources using Hive, Pig, Flume and Spark.
Prepared data sets and performed data exploratory analysis (EDA)
Performed Data pre-processing, transformation and feature engineering
Created Tree based Machine learning Model to identify fraud. Used R-Shiny and Tableau to present the claims to SIU unit for further investigation.
Used KNN and Density based models approach to identify outliers.
Performed inferential statistical analysis hypothesis testing: t-test, Analysis of Variance, Chi- Square tests and correlation on the data provided.
Performed dimensionality reduction by taking in to account interdependencies between variables and calculated R-Square and RMSE to access the Model performance.
Responsible for fine tuning, model validation and model scoring
Resolved any issues raised by the Data Engineering team during the deployment of new models.
Learned R-Shiny to provide interactive visualizations to the Business users and Operations team.

Environment: Tableau, R-Shiny, Python, R, Statistics, Machine Learning, SQL, PL/SQL, Excel, Hive, Pig, Spark, Flume

Projects and Responsibilities as Data Science Consultant for the Data and Analytics unit of Infosys (June 2014 – Till Date, Location – CT, Hartford with travel to client locations)

- Worked on multiple Infosys internal Data Analytics Projects, RFE and RFP's.
- Provided end to end Analytics and Automation solution to multiple Infosys clients across North America.
- Member of Infosys NIA(Artificial intelligence platform) team – Infosys NIA is a platform that brings together Robotic Process Automation, Predictive Automation and Cognitive Automation - An open source advanced data analytics and machine learning platform that enables businesses to operationalize their data assets and uncover new opportunities for rapid innovation and growth. As a member of Infosys NIA team, I am responsible to create Models, integrate models to the NIA platform, Model tuning, Integrate open source AI packages to the NIA platform. Use of API's
- Helped a reputed health insurance payer reduce its Late Claim Interest (LCI) cost. The solution proposed was able to predict and identify prospective claim that would result in LCI. Involved in the entire project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modeling. As part of the project clients historical claim data was ingested to the NIA platform. Used collinearity for feature selection. Created machine learning model using Naïve Bayes, Logistic regression as a classifier, SVM and Random forest. Naive Bayes was selected as the final model as it had the greatest classification accuracy. The features selected were kind of independent which means they had little correlation.
- Provided Predictive analytics tool to a plan sponsor to help predict the Health Savings Account (HSA) contribution needed by its employee's. This tool helps employees effectively manage the out of pocket expense by contributing an appropriate pretax dollar to the HSA account. The tool is highly scalable and is continuously based on the additional enrollment data that is ingested. The model is built in Python with the user interface built using JSP, JSTL, HTML, CSS, JavaScript, AJAX and JQuery technologies.
- Have experience creating word cloud using twitter data and do sentiment analysis and text analysis on the Twitter data as part of the Request to proposal for one of the Infosys clients.
- As part of consulting, Helped Infosys clients improve their claim auto adjudication rate. Identified automation opportunities and suggested an appropriate AI/ML solution.

Environment/Software: Infosys AI/ML platform – NIA, Tableau, R-Shiny, Python, R, Statistics, Machine Learning, SQL, PL/SQL, Excel, HTML, CSS, JavaScript, Hadoop, Pig.

Job Title: Technology Lead/Senior Consultant **Location: Hartford, CT**

Role: Lead Business Analyst **Client: Aetna**

June 2009 – May 2014

Responsibilities:

- Worked as a Lead Business Analyst for multiple Claims domain projects in Aetna.

- Have worked on multiple projects and requests involving Provider domain, Member domain, Pricing, network management and Clinical editing.
- Enhanced the claim processing system to implement new policies, improve claim processing turnaround time, increase claim auto-adjudication rate.
- Enhanced the claims system to identify possible fraud, waste, abuse claims and route them to the special investigation unit.
- Implemented a new overpayment system to effectively identify and track possible overpayment claims using Hadoop.
- Lead team of developers and business analysts to successfully complete multiple multi-million-dollar projects to small service requests.
- Skilled Healthcare consultant with extensive knowledge of claim adjudication system in healthcare insurance industry.
- Experienced in writing complex SQL queries like Stored Procedures, triggers, joints, and Sub queries
- Experience in Data migrations, Cleaning, Transformation, Integration, Data Imports and Data Exports as part of the projects used to migrate HMO business to strategic platform.
- Worked on mapping data from multiple systems to create a data pool and a common extract file to be used by majority of IT/Business stakeholder across the enterprise.
- Responsible for generating various reports for business team for enabling the business decisions.

Software: COBOL, JCL, DB2, Z/OS, CICS, Tableau, R, Python, SQL, HTML, JavaScript, Hadoop

---

Company: INFOSYS LIMITED, INDIA

Job Title: Software Engineer Trainee/Technology Analyst **Location: Pune, India**

Role: Application Developer **Client: Aetna**

Nov 2005 – May 2009

Responsibilities:

- Completed 6 months Induction training after joining Infosys in Nov 2005.
- Completed capstone project "Automation of Infy bank" as part of the Infosys Induction training.
- Performed Requirement and System Impact Analysis for Aetna's claim processing systems.
- Responsible for the study of SQL Queries, Analysis enhancements and documentation of the system.
- Design and Development: Prepared Design and Test plan documents and Flowcharts.
- Performed system Coding, Unit and Integration Testing.
- Implemented procedures for extracting Excel sheet data into the mainframe environment by connecting to the database using SQL.
- Used advanced Microsoft Excel functions such as pivot tables and VLOOKUP to analyze the data.
- QA testing support, Implementation support
- Lead and managed team of developers to execute the project.

- Worked as a Module lead for various projects.
- Documented claim domain learning's for knowledge management initiatives in team.

Software: COBOL, JCL, DB2, Z/OS, CICS, C, PL SQL, SAAS, Core Java, SQL, MS Excel

MAJOR ACADEMIC PROJECTS:

- **Walmart stores Time series prediction on sales per** department**.**

This Project was developed as a part of STAT 542 Practical Statistical Learning at University of Illinois, Urbana-Champaign. For this project we used the historical sales data for 45 Walmart stores that was provided to us. The aim was to build a model that can predict the Weekly Sales per store per department. The kind of problem could be posed as a time series problem with the data having temporal dependency. I implemented 3 different model in a walk Forward Validation scheme with different model combinations like STLF+SVD, TSLM+SVD, SNAIVE, ETS, ARIMA. The dataset comes from the Kaggle competition https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting

- **Sentiment analysis based on movie reviews.**

This Project was developed as a part of STAT 542 Practical Statistical Learning at University of Illinois, Urbana-Champaign. We were provided with the Movie review dataset which contains the review of a user and a binary target variable which records the sentiment of that review. We are then asked to build a model that learns to classify the reviews into one of the 2 classes. The kind of problem is a Supervised Classification Problem. The dataset comes from the Kaggle competition https://www.kaggle.com/c/word2vec-nlp-tutorial/data.

- **Housing Data Project – Predict house prices.**

As part of this project I built a predictive regression model that can estimate the price of a house using the 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa. I use Xgboost to perform the regression analysis. This data set has been used in a Kaggle competition (https://www.kaggle.com/c/house-prices-advanced-regression-techniques).

- **Lending club – Predict loan defaulters.**

As part of this project we were provided with the dataset with the historical information about its customer and how they defaulted on their loan, using all the customer attributes, I implemented a model that could classify and predict the probability of a customer defaulting on his loan. I implemented 3 models which included Random Forest, Logistic Regression and Xgboost. The dataset comes from the following Kaggle competition https://www.kaggle.com/wordsforthewise/lending-club

- **Built a Naiye Bayer classifier from scratch in R.**

Built a simple naive Bayes classifier to classify the UC Irvine machine learning data repository that hosts a famous collection of data on whether a patient has diabetes (the Pima Indians dataset), originally owned by the National Institute of Diabetes and Digestive and Kidney Diseases and donated by Vincent Sigillito. The Dataset can be found at http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes.

- **Trained SVM on the UCI Irvine Machine learning data on Adult income.**

Wrote a program to train a support vector machine on the provided data using stochastic gradient descent from scratch without using any package. The Dataset used was The UC Irvine machine learning data repository hosts a collection of data on adult income, donated by Ronny Kohavi and Barry Becker. You can find this data at https://archive.ics.uci.edu/ml/datasets/Adult

- **Performed Principal Component Analysis on the CIFAR-10 dataset.**

For the dataset provided, computed the mean image and the first 20 principal components for each category of images. Used principal coordinate analysis to make a 2D map of the means of each category. The Dataset provided was CIFAR-10 is a dataset of 32x32 images in 10 categories, collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. It is often used to evaluate machine learning algorithms. You can download this dataset from https:// www.cs.toronto.edu/~kriz/cifar.html.

- **Clustering using K-means on the European employment in 1979 dataset.**

Using k-means, clustered the dataset provided and identified the best choice of "k". Also used an agglomerative clusterer to cluster this data. Produced a dendrogram of this data for each of single link, complete link, and group average clustering. The Dataset used for dealing with European employment in 1979 can be found at http://lib.stat.cmu.edu/DASL/Stories/EuropeanJobs.html. This dataset gives the percentage of people employed in each of a set of areas in 1979 for each of a set of European countries.

- **Linear and logistic regression model with various regularizer's on the UCI machine learning dataset giving features of music and Taiwanese credit card users data.**

build a straightforward linear regression and then used Box-Cox transformation to improve it. Used glmnet to produce a regularized ridge and lasso regression. Used logistic regression to predict whether the user defaults on the credit card. Dataset used - UCI Machine Learning dataset repository hosts a dataset giving whether a Taiwanese credit card user defaults against a variety of features at http://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

- **Image segmentation using EM.**

segmented an image using a clustering method. Segmented each of the test images to 10, 20, and 50 segments. Implemented the EM algorithm from scratch (rather than using a package).Dataset used – http://luthuli.cs.uiuc.edu/~daf/courses/AML-18/EMHWPix/

- **Convolutional Neural networks** on the MNIST dataset using **Tensorflow**.

- Built a **Boltzman Machine** for denoising the images using Mean-Field Inference of the MNIST dataset consists of 60,000 images of handwritten digits.

- **Data Visualization project using Tableau, JavaScript and D3.j.**

- **End to End Data cleaning project.**

EMPLOYEMENT ELIGIBILITY: H1B Visa with I-140 approved (Priority date Feb 2013)

DECLARATION:

I hereby declare that the details stated above are authentic to the best of my knowledge and belief.