# **PANKAJ TRIPATHI**

**Data Scientist/Manager Data Science**

: 91 – 9654577826 [kontact2pankaj@rediffmail.com](mailto:kontact2pankaj@rediffmail.com)

**Keys Skills: Artificial Intelligence, Machine Learning, NLP, Statistics**

**SUMMARY:**

* MCA and MBA with 13+ years of experience in various roles across Technology and Analytics.
* Expertise in Statistical Modeling, Predictive Analytics, Machine Learning, Data Mining. Have very good experience in R and Python Programming using various packages.
* 5 years of experience in Analytics, Machine Learning assignments using R programming, Python Programming, Statistical Modeling, Predictive Analytics, Machine Learning. Successfully pulled offshore works on several occasions. Some of the highlighted works are as below -
* Predicting Mortality risk of In-patients and out patients for United Health Group, a major in US Healthcare using Logistic regression.
* Predicting the assignment group and solution for the incoming incidents in the Service desk application using SVM.
* Predicting possible delays/defaults in outstanding dues using Logistic Regression.
* Exploratory analysis for agents data who are processing healthcare claims for UHG.
* Predicting sub-functional area in the application using Naive Bayes.
* Vehicle accident report analysis using IBM Watson Machine Learning , Watson Knowledge studio and IBM BlueMix platform.
* Have worked on RODBC package to fetch data from SQL databases, have worked on R packages like dplyr, ggplot2, shiny, shinydashboards, knitr, stringr, tm, wordcloud, NLP, stats, class, gmodels, etc.
* Good Experience in text mining, cleansing, regular expressions.
* Experience in statistical modeling and techniques such as Hypothesis testing, various statistical tests.
* Good Understanding of Deep Learning – Tensorflow, Open CV, Neural Networks etc.
* Experience in visualizing data through Power BI, R- ggplot2 package, Excel.
* Extensive self-learning abilities. Hard working, enthusiast for new ideas.
* Aware of big-data concept & related tools.

**Achievements’ highlights:**

* **IBM Certified Data Scientist – Level-2**
* **IBM Certified in Deep Learning**
* **Manager's Choice Award** for Y’2016 Q4
* **Putting Client First Award** for Y’2016 Q2.
* **Orion Award** for Y’2014 for providing semi technical support to other accounts, helping in cross skilling, training and mentoring resources.

**TECHNICAL SKILLS:**

Programming Languages : R Programming, Python, JavaScript, HTML

Visualization : Power BI, Qliksense, ggplot2

Databases : SQL Server 2000-16, MS Access 2002-2013

Operating System : Windows XP/7/2000 Server/2012 Sever R2 Datacenter

Analytical skills : R Programming, Machine Learning, Exploratory Data Analysis, Classification &Regression Techniques.

Other Tools& Frameworks : IBM BlueMix, IBM Watson Analytics, Watson Knowledge Studio, Azure ML

**EXPERIENCE SUMMARY:**

**Organization Desig. Duration**

1. UHG/Optum Manager Data Science May 2018 – Present
2. IBM India Tech. Team Lead Jun 2010 – Mar 2018
3. csc India Sr. Software Engineer May 2006 – Jun 2010
4. Polaris Software Labs System Analyst Jan 2005 – Apr 2006

**EDUCATIONAL QUALIFICATION:**

* "Post Graduate Program in Management" from IMT, Ghaziabad(Part Time) in 2012 with 72% Marks.
* “Master of Computer Application (MCA)” from UPTU, Lucknow(Full Time) in 2005 with 75% Marks.
* “Bachelor of Computer Application (BCA)” from CCS University, Meerut(Full Time) in 2002 with 75% Marks.

**Certifications:**

* Completed below courses/certificates
  + IBM certified Data Scientist – Level-2
  + IBM certified in Deep Learning
  + Certificate in R Programming (Data Camp Institute)
  + Certificate in Python Programming (Data Camp Institute)
  + Introduction to Big Data (Big Data University)
  + Certified Scrum Master

**PROFESSIONAL EXPERIENCE (PROJECTS & ASSIGNMENTS):**

**UHG/Optum may 2018 – Present**

**Project: Predictive Analysis: Predicting Mortality Risk**

UHG is a major player in US Healthcare and the requirement was to predict the mortality risk for inpatients and outpatients, so that accordingly the patients can be put into the most suitable care programs to help them live a healthy and better life.

**I am able to achieve a prediction accuracy of 87% on the production/live conditions using Logistic Regression algorithm.**

**Work Environment:** RSTUDIO, Excel, dplyr, e1071, ggplot2, SQL Server.

**UHG/Optum Aug 2018 – Nov 2018**

**Project: Research study to find actionable insights from the patients risk score data.**

The objective is to conduct a research study on the patients risk scores(Future cost risk, Inpatient risk and Emergency department risk) . The objective is to find actionable insights from the data that will help management to make appropriate strategy to deal with high cost patients and also help in predicting the future high risk patients.

**Work Environment:** RSTUDIO, Excel, Power BI and SQL Server.

**UHG/Optum may 2018 – Present**

**Project: Exploratory Analysis: Predicting the performance of claims processing agents**

UHG is a major player in US Healthcare and the requirement was to predict the performance/efficiency of claims processing agents based on their daily activity data. Also find some insights that can be used by management to predict the attrition and the insights that are helpful to management to create a development plan for the team.

**I am able to achieve a prediction accuracy of 89% on the production/live conditions using XGBoost algorithm.**

**Work Environment:** RSTUDIO, Excel, dplyr, ggplot2, xgboost

**IBM INDIA LTD Jun 2010 – Mar 2018**

**Assignment (I)**

DATASCIENCETEAM: (Jul 2013 – Mar 2018

**DESCRIPTION:**

My responsibility was to provide ADHOC support for any reporting or analytical need within any of the projects/teams, being a member of COE in IBM. I have been assigned this role due to my R POCs, data exploration skills in R and my self-learning abilities.

**Project: Predictive Analysis: Resolver group prediction – Multi - Classification problem**

Project requirement is to develop a machine learning model to predict the resolver group for the incidents reported by users to the Service desk. Project involves analyzing last 2 years of data from service desk, perform text mining, cleaning and then create feature vectors. Which are then used to build a predictive algorithm to predict the multiclass(250+ target Categories).

I was able to achieve the accuracy of 78% after multiple iterations of data cleaning and extracting important text from the raw data to create better feature vectors.

**Work Environment: Python,** Regular Expressions, TF-IDF, LinearSVC.

**Project: Exploratory Analysis –HR Team**

* A new reporting solution for client’s recruitment team using exploratory analysis with R programming. My role was to find insights and propose ways to improve the overall system.
* Trends/Distribution of requirements as per type of employment, i.e. contractual and permanent.
* Role wise trends for advertisements
* Continuous support for changes required.

**Work Environment:** R Programming, RSTUDIO, dplyr, ggplot2, rodbc.

**Project: Predictive Analysis – Collection Management—Regression Problem**

Project requirement is to explore the available data from Collection Management and generate various reports. Also predict the possible delays/defaults for the outstanding dues. For this classification problem Logistic regression model was used.

**Work Environment:** R Programming, RSTUDIO, Excel,dplyr, Knitr, ggplot2, Shiny, Shinydashboard.

**Project: Sub\_Functional area prediction – Multi- Classification problem**

Project requirement is to develop a machine learning model to predict the Sub\_Functional area for a user, based on the inputs provided in Keys Skills, Job Title, Profile Title, Industry, and Experience. This was a multi-class text classification problem, which involved Data Cleansing, extraction/creation of features from Corpus, using wordcloud for visualization and so on.I have used Naïve Bayes algorithm to resolve this problem.

**Work Environment:** R Programming, RTUDIO, Excel, tm, e1071, wordcloud.

**Project: Vehicle Accident report analysis using IBM Watson**

Project requirement is to develop a machine learning model using IBM Watson Knowledge Studio and IBM Bluemix platform to analyze the Vehicle accident report analysis to assist the evaluator in the process by automating the manual work of annotations.

**Work Environment:** IBM Watson Knowledge Studio, IBM BlueMix.

**Assignment (II)**

PROJECT: e-Payables & e-purchase

CLIENT: AtLAS COPCO

**DESCRIPTION:**

E-Payables and E-Purchase applications are one of the legacy applications, developed and maintained by IBM. It is a desktop as well as web enabled application used by Accounts & Finance departments for purchases and payments, having Lotus Notes as its front-end backed by SQL server. It enhances the purchase and payment experience and enables our clients to do their tracking and reporting without any hassle.

**Responsibility:**

* My major tasks includes application support, database support and Excel reports.
* Frequent CRs I accomplished at e-payables included tasks like
  + Developing/modifying user forms/screens.
  + Developing/modifying existing code for any change in business logic.
  + Creating/modifying excel based reports.
* Database maintenance, monitoring database and server health.
* Monitoring Scheduled jobs, creating new ones when required.

**Work Environment:** HTML, JavaScript, Lotus script, SQL server 2014

**csc india** (May 2006 – Jun 2010)

sr. software engineernoida

client: codan trygg hansa & csc financial services

DEPT: insurance & Financial services

**Responsibility:**

* Provided development and maintenance support to the application portal used by CSC US Financial Services.
* Prepared different dashboards/MIS like “Bad Debt Trends”, “90 days”, “60 days”.
* Also provided development and maintenance support to the applications used by an Insurance company, Codan Trygg Hansa from Denmark.
* Extract data from SQL database through queries.
* Process automation through Macros in MS Access & MSExcel.

**Work Environment:** MS SQL, HTML, JavaScript, Lotus Notes

**polaris software labs** (Jun 2005 – Apr 2006)

systems engineer gurgaon

client: FLEXTRONICS SOFTWARE SYSTEMS

**Responsibility:**

* Provided development and maintenance support to various client's intranet applications.
* Performed developer role on the applications built on Pro\*C, PL/SQL and Lotus Notes.
* Preparing reports in MS Excel for management/client for tracking, monitoring and forecasting purpose.

**Communication Languages**: English, Hindi

**Interest/hobbies:** Travelling, Family get-togethers, computing, listening music.

Place: Noida

Date: \_ \_/\_ \_/\_ \_ \_ \_ [Pankaj Tripathi]