Renjith Paulose
Manager – Projects, Data Scientist, Big Data Analytics at Cognizant Technology Solutions

PROFILE SUMMARY_____

- 9+ years of experience in Pharma, Biotech and IT industries with deep expertise in Bioinformatics, Statistics, Data Science, Big Data, Machine Learning and Advanced Analytics solutions covering a wide range of tools and technologies.

- Deep understanding and expertise in analyzing data from diverse industries like Pharma, Biotech, Healthcare, Power & Energy, Media and Entertainment, Oil & Gas, Shipping and Logistics etc. Implemented Statistical Algorithms and Machine Learning techniques to find insights from complex datasets.

- Rich experience in Machine Learning and Predictive Modeling. Deep understanding in Statistical methods like Customer Segmentation, Data Clustering, Classification, Regression, Neural Networks.

- Expertise in R Statistical Programming, Pipeline Pilot, Weka, KNINE, R-Hadoop, Hadoop, Hive

- Technical expertise in Software Development Life Cycle entailing requirement analysis, design, development, testing and deployment.

- Good exposure in client interactions, project design and team management.

- Excellent communication, interpersonal and analytical skills with proven abilities in resolving the complex data science issues and a good team player.

EDUCATION_____

PhD. in Bioinformatics (Pursuing) – Field of Research: Machine Learning & Chemical Toxicity Prediction	**2013 - 2017**

    Bharathiar University, Coimbatore, India

    Novel Predictive Models for Toxic Effects of Chemicals in Human using Data Mining & Machine Learning

MSc, Bioinformatics	**2010**

    Sikkim Manippal University, Pune, India

Advanced Diploma, Applied Bioinformatics	**2008**

    Bioinformatics Institute of India, Noida, India

B.Tech, Industrial Biotechnology	**2006**

    St. Michael College of Engg & Tech. Anna University Madurai, India

PROFESSIONAL SUMMARY_____

Manager – Projects, Data Scientist

Nov-2014 - Till Date at Cognizant Technology Solutions, India, Kochi

Role: Data Scientist, Big Data, Advanced Data Analytics, R Analytics, Machine Learning

Bioinformatics Scientist

Jul 2012- Nov 2014 at Liatris Biosciences LLP, Kochi

Role: Data Scientist, Data Analyst, Machine Learning, Human Genome Data Analysis

Application Scientist, Principal Scientist

Sep 2008 - June 2012 at Systems Biology (I) Pvt. Ltd., Pune

Role: Bioinformatics Algorithm Developer, Data Mining, Data Clustering, Predictive Modeling, Chemical Toxicity Data Analysis

Junior Research Associate

Oct 2006 - Aug 2008 at GVK Biosciences Pvt. Ltd., Hyderabad, Chennai

Role: Predictive Modeling, Computational Drug Discovery, QSAR Modeling, Statistics

SKILLS AND TECHNIQUE_____

Data Science	Data Mining, Machine Learning, Advanced Data Analytics, Predictive Modeling, Clustering, Classification, Regression, Logistic Regression, Neural Networks, Time Series Forecasting, Algorithm Development, Quantitative and Qualitative Statistical Modeling, and Scientific Data Curation
Statistical Programming	R, Bioconductor, Weka, Pipeline Pilot, KNIME, R-Hadoop, Octave
Programming Languages	R, Perl, Python, Scala

Web Programming	HTML, PHP
Big Data	Hadoop, R-Hadoop, Hive, Pig, Spark
Operating Systems	Windows, Linux
Databases	MySQL

CERTIFICATIONS_____

- Machine Learning - Statement of Accomplishment. Stanford University through COURSERA INC. 2015
- Innovation Growth Program Certification from IC2 Institute at the University of Texas at Austin & The Federation of Indian Chambers of Commerce and Industry (FICCI) 2012
- Accelrys Certified Professional (level II) 2006

PRESENTATIONS_____

- Resource Person for 2 day National Workshop on Essentials of Bioinformatics, 2014 – URCW College – Tamil Nadu 2014
- Represented SBW in DSIR-India and Technology development Board (TDB) – India conferences. 2011-2012
- Chairperson of STOX International conference, 2009 at National Institute of Nutrition (NIN), Hyderabad for session-5: "Alternative vs. Conventional Test Systems for Toxicological Evaluation of Compounds" 2009
- Society of Toxicology, India (STOX) International conference, 2009 at National Institute of Nutrition (NIN), Hyderabad; Topic: Insilico Approaches for Toxicity Compliances. 2009
- Resource Person of GVK Biosciences for symposium; American College, Madurai;
  Topic: Recent Trends in Molecular Biology & Bioinformatics 2008

SCIENTIFIC PUBLICATIONS_____

1. Machine Learning Classifier Algorithms to Predict Endocrine Toxicity of Chemicals. Vol 7, Issue 6 - International Journal of Toxicological and Pharmacological Research – IJTPR.
2. E-Analogs: A Web-Based Cheminformatics Algorithm for Automated Ligand Analog Library Generation and In-Silico Drug Design. Vol 7, Issue 6 - International Journal of Pharmaceutical and Clinical Research – IJPCR.

PROJECTS HANDLED – REPRESENTATIVE LIST_____

Project #1

Title:	**OntoMine- Drug Discovery / Chemical Data Mining / Data Analysis Scientific Software – Novel Algorithm based on Machine Learning and Data Science**
Customer:	Pharma & Biotech – Novel R&D Scientific Software
Role:	Data Scientist / Technical Lead / Team Lead (Team Size: 6)
Duration:	3 Year 9 Months
Environment	R Programming, Weka, Pipeline Pilot statistical packages, Neural networks, Regression, Molecule fingerprints and Data Clustering

Description:

Worked as Technical lead for a large-scale Bioinformatics – Drug Discovery software development project. Several hundreds of predictive models developed and implemented using statistical methods like Clustering, Regression analysis, Neural Networks as part of this project. These Machine Learning techniques provided early stage prediction of chemical toxicity and thereby faster drug discovery.

Responsibilities:

- Hundreds of predictive models developed and implemented using statistical methods like Clustering, Regression analysis, Neural Network including Multi-Layer Perceptron (MLP).
- Technical Lead of research & development and Team Lead of Software development team.
- More than 2 Lakh unique Drug Like molecules and associated data mined, analyzed and insights generated.

Project #2

Title:	**Drug Properties Prediction Scientific Software - Novel Algorithms based on Machine Learning and Data Science**
Customer:	Pharma & Biotech - Novel R&D Scientific Software
Role:	Data Scientist / Technical Lead / Team Lead (Team Size: 5-10)
Duration:	3 Year 9 Months
Environment	R, Weka, Clustering, Regression, Neural Networks, Multilayer Perceptron

Description:

Technical Lead for Java based Scientific Tool for ADME property prediction based on Data Mining & Machine Learning. Statistical algorithms to predict Aqueous Solubility, Partition Coefficient (logP), Ionization Constant (pKa), Distribution Constant (logD) of Chemicals.

Responsibilities:

- Novel computational algorithm and models developed for chemical property prediction. Quantitative Structure Activity/Property relationship (QSAR & QSPR) using Data Science applied.
- Neural networks are used to statistically analyze and predict properties.

Project #3

Title:	**Telematics – Sales Forecasting**
Customer:	Automobile – Global Leader
Role:	Data Scientist
Duration:	1 Month
Environment	R Programming, Forecasting algorithms

Description:

Telematics Sales forecasting using statistical methods. Time series forecasting algorithms and regressions algorithms to predict future sales from historical sales data.

Responsibilities:

- Correlation among web and sales data and statistical model building.
- Time series forecasting algorithms (Forecasting sales for future years from historical data).
- Regression analysis and predictive models. Sales predictions from regression equations.
- Holt-Winters Exponential Smoothing and forecasting models.

Project #4

Title:	**Big Data – Cloudera to Amazon EMR Migration**
Customer:	Media & Publishing – Global Leader
Role:	Big Data - Project Manager
Duration:	2 Months
Environment	Hadoop, Cloudera, Amazon AWS EMR

Description:

Migrating Cloudera to Amazon AWS EMR environment. Configure existing Cloudera based MAP Reduce jobs to work on AWS EMR based Hadoop Distribution and retain the same functionality.

Responsibilities:

- Team Management.
- Project design and coordination.

Project #5

Title:	**Reader Statistics on Article Published**
Customer:	Internal POC
Role:	Big Data Analyst
Duration:	1 Month
Environment	Hadoop, Hive

Description:

Data stored on HDFS is retrieved and analyzed using HIVE. Statistics on readers like Mean number of readers, median number of readers, which day had highest number of readers etc information is calculated using HIVE.

Responsibilities:

- Develop HIVE queries for Statistical Data Analysis.

Project #6

Title:	**Human Genome Data Analysis. Bioinformatics Next Generation Sequencing Data Analysis.**
Customer:	Pharma/Biotech
Role:	Data Scientist
Duration:	4 Months
Environment	R-Programming, Bioconductor, FASTX, FASTQC, Abyss, BWA, MAQ, Bowtie, Samtools, DeSeq, Ugene, Seqmonk

Description:

Worked as Data Analyst to study human genome data generated from Next Generation Sequencing (NGS) techniques. This process involved advanced analytics using statistical methods and find perceptions from complex human genome data. Also responsible to handle day to day activities in support like Data curation, Team Monitoring, Data quality and Data Management.

Responsibilities:

- RNA Sequence data analysis on LINUX environment
- QC & processing of raw data, de novo assembly, Map reads to reference, Alignment processing, Expression, normalization and test for differences, Variant detection, annotation and filtration.
- Usage of R/Bioconductor packages like GenomicRanges, IRanges, ShortRead, Biostrings, Rsamtools, edgeR, DESeq, matplot, density plots, histograms, box plots.

Project #7

Title:	**Smart Grid Data Advanced Analytics - Insights from Smart Meter Data**
Customer:	Power and Energy
Duration:	3 Months

Role:	Data Scientist
Environment	Machine Learning, Predictive Modeling, R Programming, RStudio

Description:

Advanced Analytics focused on generating insights from Smart Meter data using Data Science.

Responsibilities:

- Customer Segmentation: Clustering Customers with Similar Usage Pattern
- Cluster Wise Data Statistics and Visualization
- Predictive Analytics: Predict Monthly Usage and Bills
- Weather - Energy usage correlations
- Time series forecasting – Total usage forecasting.

Project #8

Title:	**Twitter Sentiment Analytics – Tweets Data**
Customer:	Internal POC
Duration:	1 Month
Role:	Data Scientist
Environment	Python, R Programming

Description: Sentiment Analysis using Twitter – Tweets data crawled.

Responsibilities:

- Twitter data – Tweets crawling.
- Sentiment Analysis using statistical methods.
- Classify emotions and reporting.

Project #9

Title:	**Scientific Workflow Development for Data Analysis**
Customer:	Healthcare
Duration:	1 Year
Role:	Scientific Workflow Developer / Data Analyst
Environment	Pipeline Pilot, KNIME

Responsibilities:

- Scientific workflows developed for chemical data management and analysis using Pipeline Pilot, KNIME.
- Develop web based applications and integrate workflows to create, browse and analyze Pharma data.
- Test scientific workflows and quality reports

Project #10

Title:	**Cheminformatica Workbench – Web based Software for Chemical Toxicity Prediction**
Customer:	Pharma/Biotech
Duration:	6 Months
Role:	Bioinformatics Scientist / Developer

Environment   Machine Learning, Predictive Modeling, R Programming, RStudio, PHP

Description:

A project focused on cosmetic industry as an alternative to animal toxicity in vivo experiments. Machine Learning Algorithms implemented into Web Application for chemical hormonal perturbation prediction.

Responsibilities:

- Machine Learning and Predictive Models based on statistical analysis of chemical training data.
- Pharmacokinetics and Pharmacodynamics (PK/PD) research.
- Develop web application using PHP.

Project #11

Title: **eAnalogs – Web based Software**

Customer: Pharma/Biotech

Duration: 6 Months

Role: Bioinformatics Scientist, Developer

Description:

eAnalogs is a novel algorithm and web-accessible software for automated ligand analog library generation for PHARMA/BIOTECH.

Responsibilities:

- Develop novel algorithm for automated combinatorial analog library construction from single scaffold.
- PHP, Perl, server configuration for Web-based software development and deployment

Project #12

Title: **Drug Toxicity Database Curation and Data Analytics – Personalized Medicine Research.**

Customer: Healthcare Consortium

Duration: 8 Months

Role: Bioinformatics Scientist / Project Manager / Team Lead (Team Size: 12)

Environment   MS Office, R programming

Description:

Toxicity data of drugs and chemicals carefully mined and curated for advanced data analytics. Patient heathcare and genetic data is analyzed for personalized medicine research.

Responsibilities:

- Scientific Data Curation Manager for Chemical/Drug Toxicity Database curation/maintenance team.
- To manage team of biologists and chemists in order to curate and develop a drug toxicity database.
- Large drug toxicity data is statistically analyzed, inferences and graphical reports generated using R and other packages
- Client interactions and project design and management.

Project #13

Title: **Computer Aided Drug Design (CADD) and Predictive Modeling**

Customer: Pharma/Biotech

Duration: 6 Months

Role: Team Member / Researcher

Environment Accelrys software tools, Cerius2, Insight II, Catalyst, Discovery Studio

Responsibilities:

- Structure Based Drug Design (SBDD), Analogue Based Drug Design (ABDD) using Accelrys Cerius 2, Insight II, Catalyst, Discovery Studio., Software Used: ISIS Base, ISIS draw, ACD, ChemOffice.
- Quantitative and Qualitative modeling for drug property prediction. Statistical Analysis and chemical property prediction.
- Molecular modeling, Molecular Docking, Protein Modeling, Force field, energy calculation energy minimization studies, Denovo ligand design and optimization, Pharmacophone design, virtual high throughput screening (vHTS)
- Molecular Mechanics, Molecular Dynamics, Quantum Mechanics Researches.