**Chun Ma**

**Green Card holder | Do not need sponsorship**

machun314@gmail.com | Los Angeles, CA, 90803| 6198001596| www.linkedin.com/in/chun-ma-734963110

**PROFESSIONAL SUMMARY**

* Results driven and self-motivated data scientist and programmer with 7 years of hands-on work experience.
* Expert in advanced statistical analysis, database, machine learning, programming and modeling.
* Expert in providing valuable insights from big size data and presenting to non-technical business audiences.
* Expert in identifying business values and developing efficient solutions to improve business success.
* Expert in driving business effectiveness through computer science and data science applications.

**MAJOR SKILLS**

Programming (R in R Studio, Python, SAS), Database (SQL,SSIS,Spark), Mathematics and Statistics, Machine Learning and Cloud Computing(Microsoft Azure, AWS), Predictive Modeling, Communication and Interpersonal skills, Data Visualization (ggplot2, Matplotlib, Tableau, GIS), Organization, Management and Leadership skills.

**EDUCATION**

Ph.D. Modeling and Data Integration in Environment Science 2011

* + University of Wisconsin- Madison and Nankai University Joint Program

M.S. Data Analysis in Environment Science 2008

* + Nankai University

**EXPERIENCE**

**Molina Healthcare, Inc., Long Beach, CA 03/2017 to present**

**Data Analyst and Programmer on Predictive modeling**

* Developed time series predictive models (which is supervised learning in machine learning) by using Tseries, forecast, ggplot2, broom, TSA etc. packages in R with monthly, weekly retrospective and projective data for STAR projections for HEDIS (which are healthcare quality measurement tools).
* Evaluated the best model by comparing Arima, AutoArima, Sarima for time series model and automated the whole process by using flashdashboard, Rmarkdown, Rbookddown, R sweave in R for reproducible work.
* Visualized and operationalized R data in Power BI for time series analysis.
* Expertise in R programming, R Studio interface, R packages such as Flexdashboard, Shiny, rattle, TWANG package, taskscheduleR, AppliedPredictiveModeling, knitr, tm packages, etc.
* Conducted data mining, cluster, classification and regression tree (CART), and random forests with a graphical user interface Rattle package in R.
* Converted SAS code to R code, and transferred data to R.
* Developed HEDIS/CAHPS related reports using SQL Server Reporting Services.

**San Diego State University, San Diego, CA 04/2016 to 01/2017**

**Research Associate on Modeling and Data Integration**

* Conducted data preparation, data cleaning, data exploring, data mining, data visualization, outlier detection, variables selection with SQL, R programming, and Python.
* Performed statistical analysis, statistical predictive modeling, and other mathematical modeling with R and Python, lead tasks in activities related to data and mathematical model integration.
* Converting large scale data sets, and dealing with data from satellites, sensors, and various data products.
* Built simple liner regression model (SLR) and multiple linear regression model (MLR) and validate the models.
* Expanded business by presentations, reports and visualized data for technical and non-technical audiences.

**The Henry M. Paulson Jr. Institute, Chicago, IL 01/2015-07/2015**

**Data Scientist on InVEST Model simulation**

* Implemented data manipulation by SQL, and data cleaning and preparation, and performed variable selection and model validation by R in R studio and InVEST model.
* Performed descriptive statistics, exploratory analysis, linear regression, non-linear regression, time-series analysis and forecasting with R by using dplyr and TSA packages.
* Interpreted data from primary and secondary sources using statistical techniques and provided ongoing reports.
* Worked with internal and external partners to understand project requirements and strategy, and support those goals with modeling and analytics
* Communicated results through presentations, publications, reports at international conferences.

**Carnegie Institution for Science at Stanford University, Stanford, CA 12/2013 – 02/2016**

**Visiting Scholar on Data Science for Ecosystems Analysis**

* Extracted knowledge and insights from data on remote server, query data from SQL database.
* Built linear and non-liner models and conducted data classification, regression, clustering and visualization with R programming.
* Conducted analysis based on mathematics and statistics such as frequency distribution, analysis of variance (ANOVA), correlation techniques, and factor analysis.
* Implemented the machine learning algorithm to analyze ecosystems, build, test and validate model.
* Collaborated with other experts on the interdisciplinary team to solve problems using data science approaches and techniques.

**Ministry of Transport, China 07/2011 – 12/2014**

**Data Scientist on Environmental Analysis Systems**

* Developed and carried forward a coherent research strategy in predictive modeling, statistical learning and environmental information systems.
* Extracted, compiled, and analyzed data to generate reports by R and Python.
* Supervised and mentored junior scientists in research projects related to PCA, CCA, linear and non-linear model, and decision tree; test, validation, and evaluation of models.
* Managed databases and data operation team and initiated connections with internal departments and centers, external institutions, universities and professional organizations to actively foster collaboration.
* Presented and communicated the data insights in publications.

**Nankai University 07/2008 – 07/2009**

**Research Assistant on Environmental Data Science**

* Conducted innovative data-driven research into complex scientific problems, ideas, concepts and theories and applied predictive modeling and other appropriate mathematics-based methodologies with SAS, SQL and ArcGIS.
* Worked with other team members to complete special projects and achieve project deadlines.
* Conducted descriptive statistics, linear regression, nonlinear regression and predictive analysis with SAS and R programming.

**Nankai University 09/2006 – 08/2007**

**Research Assistant on Data Analysis**

* Implemented statistical analysis of retrieved data, experimental data and model output data with SAS and R.
* Conducted data review and data cleaning of raw data.
* Performed daily data queries and prepared reports on daily, weekly, monthly, and quarterly basis.

**ACCOMPLISHMENTS**

* 33 peer-reviewed papers based on data analysis and modeling; 12 data-driven research projects;
* 23 presentations on international conferences; Chapters for 7 books on science and engineering.

**CONTACT INFORMATION**

machun314@gmail.com | Los Angeles, CA, 90803 | 6198001596| www.linkedin.com/in/chun-ma-734963110