Peng Chen

Data Scientist | Consultant

814 810 8589 | pfc5098@gmail.com | pfc5098 | pfc5098

Skills

- Advanced R
  - Data engineering: DBI, glue (for SQL abstraction), dbplyr, dplyr, dtplyr, data.table, arrow, SparkR, sparklyr
  - Data exploration and communication: ggplot2, highcharter, gt, DT, rmarkdown, flexdashboard, tidyverse
  - Predictive modeling / machine learning: tidymodels, torch, reticulate, tensorflow, keras, h2o, rsparkling
  - Statistical inference and modeling: infer, rstanarm, rjags
  - Natural Language Processing (NLP): tidytext, textrecipes
  - Network analysis: igraph, ggraph, visNetwork
  - High performance computing: parallel, doParallel, doMC, foreach, Rfast
  - Package/APP/API development and productization: devtools, shiny, shinydashboard, plumber, pins, HTML/CSS
  - Repository and library management
- Spark/Databricks/PostgreSQL - logical execution order, temp tables, CTEs, array, struct, window functions, stored procedures
- Python - pandas, scikit-learn, PySpark, Boto3
- Distributed computing frameworks for big data - Apache Spark, H2O
- Cloud Services - AWS (EC2, S3, RDS, Redshift, Rekognition, Translate, Comprehend), Databricks (DBFS, Delta Lake), RStudio Server
- Bash, PowerShell, Git, GitHub, Bitbucket

Work Experience

Axtria - Ingenious Insights

Berkeley Heights, NJ

Decision Science Senior Associate Consultant

2020 - Present

- Successfully defined the market of a new brand for Pharma A by coordinating with teams across different organizations and time zones.
- Secured a new brand supporting project for Axtria from Pharma A by demonstrating extensive analytical skills and experiences.
- Led payer-stream consulting efforts for Pharma B. Timely responded to big volumes of client requests, effectively communicated with clients to understand underlying business problems, proactively cleared roadblocks and assigned tasks to offshore colleagues to form a seamless global workflow, quickly took feedback from clients and iterated, and delivered satisfactory results either on time or ahead of time.
- Kept team members working effectively with rational workload-control techniques. Broke client requests into actionable sub-tasks and prioritized them by importance and deadlines. Released only bottleneck sub-tasks to only team members who had bandwidths. Managed client expectations and re-prioritized ongoing requests when the team was working at its full capacity, protecting their mental health and creativity, resulting in high-quality deliverables that clients appreciated.
- Collaborated with Axtria teams and successfully delivered a Physicians & Hospitals Segmentation consulting project to Pharma B. Learned all aspects of Life Sciences commercial segmentation practices from Axtria team members and clients.
- Worked on a wide variety of Life Sciences projects: doctors and hospitals segmentation and targeting, promotional response modeling, marketing mix optimization, patient analytics such as patient journey analysis and potential patients identification, market access analytics such as payer-plan segmentation, contracting and rebate optimization, payer favorability modeling, etc.
- Internal consulting: sub-national forecasting, doctor market access scoring, virtual account manager allocation, RStudio Server setup, etc.
- Recommended learning resources on data science, AIML, and software engineering to cultivate a data-centric culture within Axtria.
- Extensive knowledge of commercial Life Sciences data sets, such as IQVIA NSP, DDD, Xponent, Xponent PlanTrak, NPA, LAAD, ELAAD, Veeva promotional data sets, etc.

Axtria - Ingenious Insights

Berkeley Heights, NJ

Decision Science Senior Associate Researcher and Developer

2020 - Present

- Innovated a lookalike model to make it much faster to run, easier to interpret and yield better prediction performance all at the same time. This new method has been well appreciated and adopted by the Axtria team and leadership. It is being used to tackle problems across patient analytics, physician analytics, and hospital analytics.
- Solved the long-lasting memory issue of an AI/ML-based patients-lookalike process by implementing a generic map-reduce strategy in R. (Patient-claim data is too big to fit into a 2TB RAM. Moreover, only 30% of RAM can be used to ensure the server's high performance.)
- End-to-end automated several payer-stream standard reports for Pharma B using R Shiny. For some of the reports, the refreshing time has been reduced from hours to minutes. Other reports even begin to "breathe" on their own such that no more human intervention will ever be needed. Our clients told us they loved these reports because they are not only automated but also improved with more concise and cleaner KPI representations. These positive changes enabled our clients to make quicker and smarter decisions at the same time.
- Developed a set of R packages for Pharma B, facilitating new members to pick up work quickly. For example, these packages enabled us to connect to databases, pull up commonly used SQL snippets, refresh standard reports, generate beautiful dashboards, and make use of third-party APIs with single lines of code. These packages have been helping us stay away from repetitive tasks and focus on more creative ones.
- Continuously improve the performance and user experience of a payer-contracting software, written in R Shiny and SQL, by adding custom R packages as patches, refining the underlying infrastructure, and introducing new features.
- Developed a NLP application in R to help Pharma C's HR department identify risky items.

Education

The Pennsylvania State University *University Park, PA*

PhD major in Industrial Engineering & Operations Research (IEOR) and minors in Statistics and Mathematics *2015 - 2020*

- Advisor: Enrique Del Castillo, Distinguished Professor of Industrial Engineering and Professor of Statistics
- Dissertation: Confidence Regions for Response Surface Optima Based on Parametric and Nonparametric Models

University of Missouri - Columbia *Columbia, MO*

MS in Industrial Engineering *2013 - 2015*

East China University of Science and Technology *Shanghai, China*

BS in Electrical Engineering, BA in Financial Management *2009 - 2013*

Academic Research

The Pennsylvania State University *University Park, PA*

Graduate Research Assistant *2016 - 2020*

- Developed a set of algorithms to compute nonparametric and distribution-free confidence regions on the optima of multi-dimension functions fitted from data, and (non)parametric Bayesian credible regions on the same. The accompanying R package, OptimaRegion, has been used by nearly 20,000 users worldwide to this date. These algorithms also have the potential to systematically solve the hyperparameter tunning problem in the machine learning field.
- Investigated the United States IEOR faculty hiring network with a latent variable exponential random graph model.
- Applied statistical process control and shape analysis to point-cloud data to improve the quality of 3D-Printing products.

University of Missouri - Columbia *Columbia, MO*

Graduate Researcher *2014 - 2015*

- Formulated the Toyota Kanban production system as a Markov Chain and maximized its profits using the Simulated Annealing Algorithm.
- Built a sales transaction database for a medical equipment company in Shanghai, China.

Teaching

The Pennsylvania State University *University Park, PA*

Instructor of Stochastic Models in Operations Research (IE 425) *2019 - 2020*

- Independently developed lectures, computer simulations, homework assignments, and exams.
- Worked with a teaching assistant and a grader to serve 132 undergraduate students of this core IE course.
- Learned how to explain things in the easiest way possible.

The Pennsylvania State University *University Park, PA*

Teaching Assistant of Engineering Economics (IE 302) *2015 - 2016*

- Helped the instructor to teach cash flow analysis and project valuations.
- Held office hours and review sessions, and graded homework assignments and exams.

Sample Courses

- Mathematics: linear algebra, real analysis, functional analysis, numerical computations at Penn State University (PSU)
- Optimization: linear programming, convex optimization, nonlinear programming, statistical process optimization at PSU
- Statistics: linear models, experimental design, probability theory, statistical theory, stochastic process and MCMC, spatial statistics, time series at PSU
- Artificial Intelligence and Machine Learning: machine learning at PSU, deep learning specialization at deeplearning.ai

Selected Publications

- Chen, P., & Castillo, E. del. (2021). On inferences about the location of response surface optima based on a latent spatial process (under revision). Journal of Statistical Computation and Simulation.
- Castillo, E. del, Chen, P., Meyers, A., Hunt, J., & Rapkin, J. (2020). Confidence regions for the location of response surface optima: The r package OptimaRegion. Communications in Statistics-Simulation and Computation, 1–21.
- Castillo, E. del, Chen, P., Meyers, A., Hunt, J., & Rapkin, J. (2019). OptimaRegion: Confidence regions for optima. CRAN.R-project.org/package=OptimaRegion.
- Castillo, E. del, Meyers, A., & Chen, P. (2020). Exponential random graph modeling of a faculty hiring network:The IEOR case. IISE Transactions, 52(1), 43–60.