|  |
| --- |
| **[Werner Chao](https://www.linkedin.com/in/wernerchao/)**  <https://github.com/wernerchao> **|** [236-777-8852](tel:2367778852) (Toronto) **|** Skype: [wernerchao](skype:wernerchao) **|** [werner.chao@gmail.com](mailto:werner.chao@gmail.com?subject=RE:%20Received%20Your%20Resume)  Blog: <https://blog.nycdatascience.com/author/wernerchao/> |

|  |  |
| --- | --- |
| **SUMMARY**  Data scientist with 4 years of experience in health tech, consumer electronics, and construction sectors. Proficient in using Python for data mining, supervised, & unsupervised machine learning. Experience analyzing big data with Spark, Hadoop, & MapReduce. | |
| **TECHNICAL SKILLS** | |
| **Programming Languages:** | Python (tensorflow, pandas, numpy, scipy, matplotlib, scikit-learn, scrapy)  R (dplyr, ggplot2, shiny, caret) **|** SQL **|** NOSQL(MongoDB) **|** JSON **|** MatLab **|** Javascript  Spark (PySpark, MLlib) **|** Hive **|** Hadoop **|** MapReduce |
| **Machine Learning:** | XGBoost **|** Random Forest **|** K-Means **|** Naïve Bayes **|** GLM **|** ARIMA **|** SVM **|** SVC  NeuralNetwork **|** Natural Language Processing |

|  |  |
| --- | --- |
| **EDUCATION** | |
| **NYC Data Science Academy**  **Data Scientist**  A highly intense & immersive data science program providing hands-on experience with machine learning, big data, R & Python development, deep learning, Hadoop, Spark, Unix, SQL, & Git | **2017.2 – Present** |
| **University of Toronto**  **Master of Materials Science & Engineering** **(CGPA 3.85/4.0)**  Thesis: Analysis of solar cells efficiency enhancement with nanopatterns **(Grade: A)** | **2011.9 – 2014.1** |
| **Bachelor of Materials Science & Engineering** **(Dean’s Honor Graduate)** | **2006.9 – 2011.5** |

**PROJECT HIGHLIGHTS**

House Price, Advanced Regression – Predicted house prices in Iowa **(top 10% in competition, 2000 participants)**

* Boosted RMSE accuracy by 13% by combining ***5 predictive models*** (lasso, ridge gradient boosting, knn, & linear regression)

Allstate Insurance Claim Severity Regression Prediction – Predicted insurance cost for car accidents **(top 10% score)**

* Boosted MAE accuracy by 7% using a master model by combining 3 different prediction models ([blog post link](http://blog.nycdatascience.com/student-works/improving-model-accuracy-kaggle-competition/))
* ***Predictive Models***: gradient boosting trees, multilayer perceptron, linear regression, lasso / ridge regression & random forest

Restaurant Sales Prediction – Conducted sales forecast for 2 restaurant locations using time series models

* Used ***ARIMA time series model*** and ***XGBoost*** for sales forecast. Improved baseline model by 80%

Conducted in-depth analysis of [NBA statistics](https://blog.nycdatascience.com/student-works/nba-data-exploratory-3-point-shooting-just-hype/) and predicted team’s offensive rating

* Scraped and analyzed NBA data using scrapy and statistical testing methods ([blog post](https://blog.nycdatascience.com/student-works/scraping-nba-statistics/) link)
* Built data visualization with R([Shiny App](https://chaowerner.shinyapps.io/shiny_proj/), ggplot2 & dplyr), predicted offensive rating using ***GLM model***

|  |  |
| --- | --- |
| **PROFESSIONAL EXPERIENCE**  **Data Analyst, KaJin Health (www.kajinonline.com) – Online Mental Health Marketplace 2016.2 – 2016.12**  Supported product & business development team using analytics, including:   * Increased sales efficiency by 20% & identified specific active market segment using cluster analysis for market segmentation * Conducted A/B testing; analyzed customer, traffic, keywords, & trend data to lead product development | |
| **Research Scientist, Appetite Lab**  Researched, developed, & launched a new fishing product currently sold on Amazon & Walmart:   * Fixed signal & buoyancy issues by analyzing 100MB research data with MatLab * Optimized design geometry through computer simulation. Conducted lab & field testing of prototypes | **2014.8 – 2015.8** |
| **Data Analyst, Project Management SNC-Lavalin 2014.2 – 2014.9**  Managed Vale’s multi-billion-dollar copper mine project in Sudbury, including:   * Programmed in VBA to analyze cost variations, cash flow, cost trends, & visualized in a storytelling manner * Forecasted monthly & annual budget based on historical economics. Reported to clients & VPs in corporate standard | |

|  |
| --- |
| **HIGHLIGHTED ACHIEVEMENTS**   * Rank ***top 10%*** in the Kaggle House Price Prediction competition, out of 2,000 participants * Awarded Ontario Graduate Scholarship, NSERC Scholarship, & Materials Innovation James Toguri Scholarship |