CINDY HUANG

Data Scientist Intern

- San Jose, CA
- +1 (919) - 816 - 5199
- yuanhsin8311@gmail.com
- github.com/yuanhsin8311

SKILLs

- **Scripting Languages:** Python, Unix Shell Scripting, Scala
- **Database Languages:** MySQL, Oracle SQL, PL/SQL, T-SQL
- **NoSQL Database:** MongoDB, Neo4j, Cassandra
- **Big Data Analytics Tools:** Apache Hadoop (MapReduce, Sqoop, Flume, Pig, Hive, Impala, Oozie), Apache Spark (SQL, MLlib, Streaming, GraphX), Google BigQuery
- **AWS Tools:** S3, EC2, EMR, RDS, Aurora, RedShift, Data Pipelines
- **ETL Tools:** SSDT-BI (SSIS), Talend
- **Statistical Tools:** R, SAS
- **Reporting Tools:** Tableau, SSRS, Google Data Studio
- **Web Languages:** HTML, CSS, JavaScript, Python- Flask
- **Languages:** Mandarin (Native)

CERTIFICATE

- First Class in Java   2016/05
  UC Berkeley Extension
- Data Scientist   2016/09 - 2017/01
  Data Application Lab

HONOR

- **IBM Big Data for Social Good Challenge**
  Grand Prize (2015/04)
  http://goo.gl/mYKZC1

EXPERIENCE

2016/10 - 2017/01

Data Scientist Intern
Neutron Mobile, California USA

- **Established a In-Game Recommender Engine**
  - Used Python (Requests, BeautifulSoup) to crawl Steam user data through Steam Web API.
  - Adopted Python Pandas to process data and build the MySQL database via Python SQLAlchemy.
  - Ran PySpark MLlib on Amazon EMR to deploy a recommender engine.
  - Implemented the testing environment using Python-Flask.

2010/01 - 2010/10

Senior Product Marketer (Retention Team)
Taiwan Mobile Co. ,Ltd, Taiwan

- **Sales Management (Unlimited Data Plan)**
  - Established monthly sale plans for various sales pipeline including direct sales, franchise, e-commerce and telemarketing.
  - Scheduled the campaign automation workflows via SAP CRM.
  - Created KPI dashboards to track weekly sales performance and make ad-hoc analysis to identify the sales risk or opportunity using Excel Pivot Tables.
- **Marketing Analytics**
  - Developed subscriber churn prediction and five-years sales prediction using SAP BusinessObjects Enterprise.
  - Derived Google Analytics on conducting A/B split testing to improve click-throughs and engagement on the device promotion banner.

2008/06 - 2010/01

Product Marketer (Category Team)
AVON Products Inc. , Taiwan

- **Modelling Database Design and Demand Forecasting**
  - Applied Talend to clean data from different teams through executing ETL process and examine the data quality.
  - Rebuilt database from EER diagram to integrate all product categories' historical sales data using MySQL.
  - Leveraged SAS/ACCESS on deploying the multiple linear regression predictive model to mitigate sales risk.
- **Product Management (Fragrance)**
  - Managed the product life cycle and optimized the size of line portfolio.
  - Executed 360-degree activation plan for quarterly newness.
  - Designed dashboards of ROI/Portfolio Analysis to monitor campaigns.

2004/07 - 2005/11

Buyer
Y F Chemical CORP. , Taiwan

  - Constructed monitoring dashboards to do suppliers' assessment.
  - Sourced new suppliers and collaborated with strategic suppliers.

EDUCATION

2013 - 2015

MS, Computer Information Technology
Purdue University, Indiana USA
Thesis: Video Advertisement Mining for predicting revenue using Random Forest

2006 - 2008

MBA, Marketing
National Central University, Taiwan
Thesis: The Study on Logistics Service Quality in B2C E-Commerce

2000 - 2004

BA, Russian
National ChengChi University, Taiwan

PORTFOLIO

- **Analytical Projects (Python / R)**

  (a) Python (Packages: Numpy, Pandas, Matplotlib, Seaborn, Sklearn)

  1. UBER Rider Churn Prediction using Logistic Regression

  2. Customer Purchase Behavior Prediction using Random Forest

  3. Rental Listing Popularity Prediction using Trees (Random Forest/Gradient Boost)

  (b) R (Packages: dplyr, ggplot2, plotly, quantmod, tseries)

  1. Coupon Purchase Prediction using Cosine Similarity

  2. Stock Revenue Prediction using Time Series Models (ARIMA and GARCH models)

- **Big Data Projects (Hadoop / Spark)**
  1. To Design a Data Warehouse for a Retail Shop using Hadoop Tools and Spark
  2. Twitter-Based Job Portal Service using Hadoop Tools

  3. Processing Movielens Dataset using Hadoop Tools on AWS Platform

  4. Querying Airline On-Time Performance Dataset using Hive

  5. Data Ingestion between RDBMS and HDFS using Sqoop

- **Data Visualization (Tableau / R)**

  1. Using Tableau to analyze Airbnb Listing

  2. Processing NYSE Trading Data using Hive and Tableau

  3. To Visualize Crime Analysis using Hive, Impala and R