Danny Ki [Data Scientist]

CONTACT

Address
139 Fairway Dr.
Newnan, GA 30265

Mobile
(678)-850-4240

Email
kish1919@gmail.com

Website
datavoyagerdanny.com

GitHub
github.com/kish191919

Linkedin
Linkedin.com/in/danny-sunghwan-ki-52381116

SKILLS

Data Science[Python]
Numpy/Pandas.
Sklearn/Statsmodel
Seaborn/ggplot

Languages
Python
R

RDBMS and NoSQL
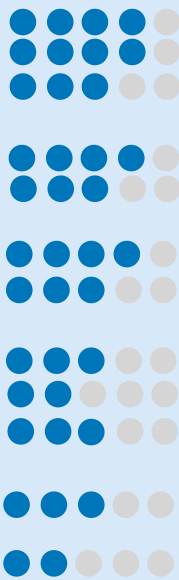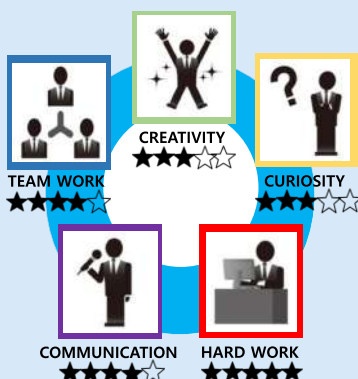MySQL
MongoDB

Big Data
Hadoop
Spark
Hive

Framework
Flask

Cloud
AWS

PERSONAL

CREATIVITY ★★★☆☆
TEAM WORK ★★★★☆
CURIOSITY ★★★☆☆
COMMUNICATION ★★★★☆
HARD WORK ★★★★★

EDUCATION

- **FAST CAMPUS** - Seoul, Korea   JAN 2018 - MAY 2018
  - Data Science School

- **UDACITY** - Online Course   JUL 2017 - AUG 2017
  - Intro to Programming Nano degree

- **MYOUNGJI UNIVERSITY** - Seoul, Korea   MAR 2000 - FEB 2009
  - Bachelor's Degree in Business Administration and International Business

PROJECT   Portfolio Website: http://datavoyagerdanny.com

- **Predict Used-Car Price in Georgia [Service Website : http://dannyki.ga/] / XGBoost**
  - Crawled on cars.com to collect data and store it in AWS's mysql.
  - After preprocessing stored data, put it in machine learning model to predict used car price.
  - Models learned on AWS server implemented as web services using Flask web framework.

- **[KAGGLE Competition] Predict House Prices / OLS Regression**
  - Developed and applied OLS algorithms to predict house prices in Ames, Iowa
  - It was the first project I submitted to the Kaggle Competition and solved the problem with a probabilistic approach.

- **[KAGGLE Competition] Spooky Author Identification / Naive Bayes Classification**
  - As a text analysis project, it was a problem seeing which authors wrote articles. After vectorizing words, classified them via machine learning with Naive Bayes Classification .

- **[KAGGLE Competition] Titanic Machine Learning from Disaster/ Voting Classifier**
  - Predicted survival on the Titanic.

- **[KAGGLE Competition] Bike Sharing Demand / Random Forest**
  - Forecasted bicycle demand using R language.

CERTIFICATION

FAST CAMPUS
- Machine Learning with R
- Apache Hadoop

COURSERA
- Machine Learning *(Andrew Ng)*
- Machine Learning Foundation *(Carlos, Emily)*
- Introduction to Probability and data *(Mine)*
- Python Programming *(Charles Severance)*

EXPERIENCE

Sales and Logistics Assistant Manager   AUG 2015 - DEC 2017

Dongwon Autopart Technology Georgia LLC (web site)   *Hogansville, Georgia*
- Analyzed financial reports, markets, and other data to maximize profit and minimize costs
- Established the negotiating strategy for sales and took the lead position in automotive markets

Purchasing Assistant Manager   DEC 2013 - JUL 2015

Kukdo Chemical Co.,LTD. (web site)   *Seoul, Korea*
- Negotiated with various vendors to ensure that a fair price for goods using ICIS and Platts
- Monitored and evaluated supplier metrics for capability, performance, and delivery to meet organization planning and forecasting needs.

Purchasing Specialist   OCT 2011 - NOV 2013

Lotte Chemical Alabama Corp (web site)   *Auburn, Alabama*
- Planned and purchased chemical materials from domestic and imported vendors with economic conditions.

PROJECT

(1) Predict Used-Car Price in Georgia

Subject : Machine learning based vehicle forecasting program
Period : 2018. 03 - 2018. 04
Tech : Python (Pandas, Scikit-learn), Data Crawl , AWS, Flask, MySQL, Bootstrap
Model : XGBooster (Accuracy : 85%)
Structure :
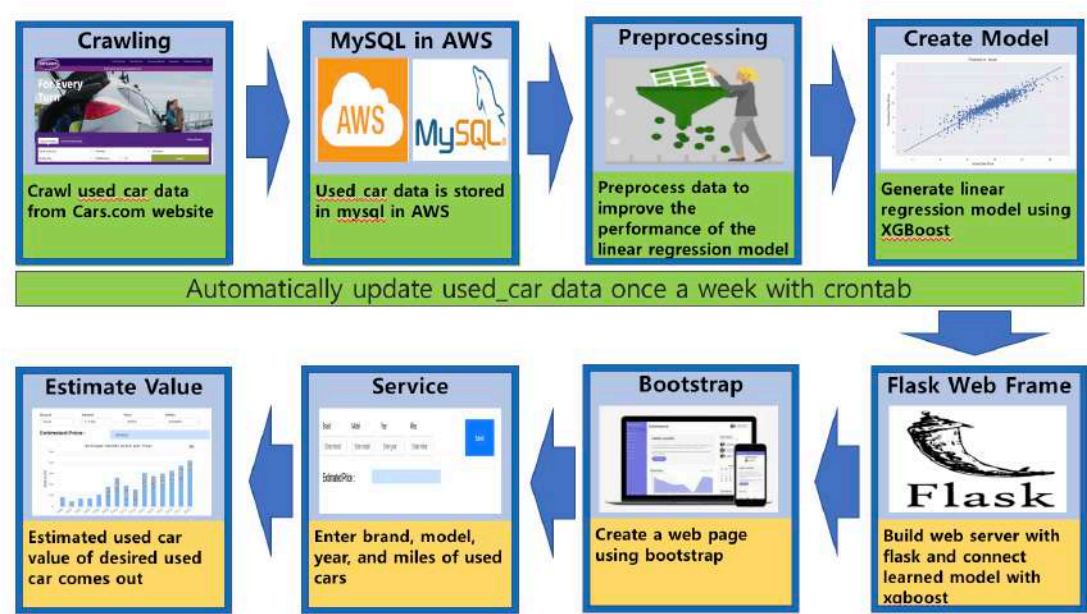
Service Website : http://dannyki.ga/
How to use the Service Website :
➜ Fill in the information and then press the submit button

➜ You can check the price of the used car you want, and you can also check the average price for different years with the same model
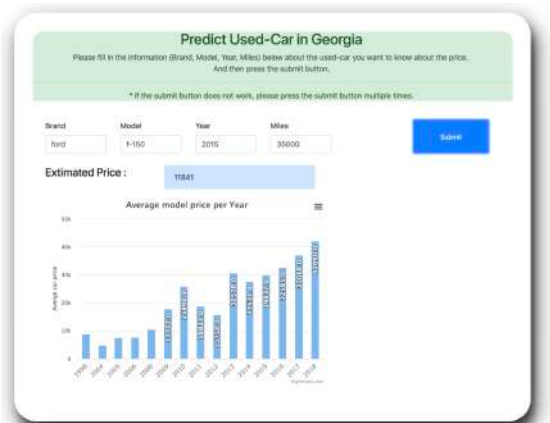
PROJECT

(2) [KAGGLE Competition] Predict House Prices

Subject : Predict house prices in Ames, Lowa
Period : 2018. 01 - 2018. 03
Data : Train Data - 81 variables and 1460 house data
          Test Data - 80 variables and 1459 house data
Python : Preprocessing - Numpy, Pandas
         Graph - Matplotlib, Seaborn
Model : **Ordinary Least Squares Model**

```
                          OLS Regression Results
================================================================================
Dep. Variable:              SalePrice   R-squared:                      0.945
Model:                            OLS   Adj. R-squared:                 0.942
Method:                 Least Squares   F-statistic:                    403.6
Date:                Mon, 26 Mar 2018   Prob (F-statistic):              0.00
Time:                        21:13:50   Log-Likelihood:                1405.0
No. Observations:                1383   AIC:                           -2696.
Df Residuals:                    1326   BIC:                           -2398.
Df Model:                          56
Covariance Type:            nonrobust
================================================================================
                          coef    std err       t      P>|t|    [0.025   0.975]
--------------------------------------------------------------------------------
Intercept                12.1018    0.059   204.753    0.000    11.986   12.218
C(Neighborhood)[T.Blueste] -0.0265  0.069    -0.381    0.703    -0.163    0.110
C(Neighborhood)[T.BrDale]  -0.0495  0.037    -1.347    0.178    -0.122    0.023
C(Neighborhood)[T.BrkSide] -0.0058  0.031    -0.184    0.854    -0.067    0.056
C(Neighborhood)[T.ClearCr] -0.0454  0.033    -1.383    0.167    -0.110    0.019
```

Kaggle Score : 0.12384 / Kaggle rank : 1042 / 4548 (22.9%)
Github : https://github.com/kish191919/House_Price_Project_by_Python

(3) [KAGGLE Competition] Spooky Author Identification

Subject : Identify an author from sentences which they wrote
Period : 2018. 03 - 2018. 04
Data : Train Data - 3 variables and 19,579 text data
        Test Data - 2 variables and 8,392 text data
Python : Natural Language Processing - Stopword, Stemming
        Vectorization - CountVectorizer
        Model - Randomforest, AdaBoost, SVM, Naive Bayes Classification
Model : **Naive Bayes classification**

```
Confusion Matrix :
 [[7414  110  376]
 [ 631 4764  240]
 [ 588   89 5367]]

10-fold Cross Validation Report:
            precision    recall   f1-score    support

         0      0.86       0.94       0.90        7900
         1      0.96       0.85       0.90        5635
         2      0.90       0.89       0.89        6044

avg / total     0.90       0.90       0.90       19579
```

Kaggle Score : 0.48767 / Kaggle rank : 793 / 1244 (63.7%)
Github : https://github.com/kish191919/Spooky_Author_Identification_by_Python

(4) [KAGGLE Competition] Titanic Machine Learning from Disaster

Subject : Predict survival on the Titanic
Period : 2018. 03 - 2018. 04
Data : Train Data - 12 variables and 891 data
　　　　　Test Data - 11 variables and 418 data
Python : Preprocessing - Numpy, Pandas
　　　　　Graph - Matplotlib, Seaborn
　　　　　Models - DecisionTree, Randomforest, Adaboost, Support Vector Machine,
　　　　　　　　　Naive Bayes Classfication, VotingClassifier
Model : **VotingClassifier Model**

```
Confusion Matrix :
 [[484  57]
  [ 80 260]]


10-fold Cross Validation Report:
            precision    recall   f1-score    support

        0       0.86      0.89       0.88        541
        1       0.82      0.76       0.79        340

avg / total     0.84      0.84       0.84        881
```

Kaggle Score : 0.78468 / Kaggle rank : 4304 / 10676 (40.3%)
Github : https://github.com/kish191919/Titanic_Machine_Learning_from_Disaster_by_Python

(5) [KAGGLE Competition] Bike Sharing Demand

Subject : Predict demand on bike
Period : 2018. 04
Data :Train Data - 12 variables and 10,886 data
　　　　　Test Data - 9 variables and 6,493 data
R : Preprocessing - dplyr
　　　　　Graph - ggplot
　　　　　Model - Randomforest
Model : **Randomforest**
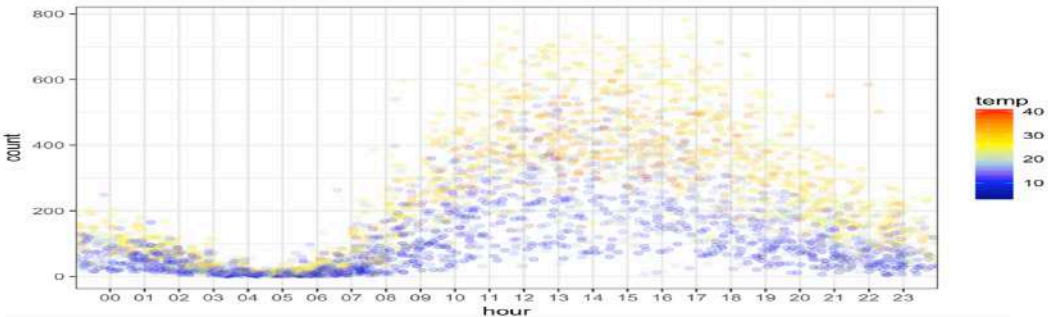
Kaggle Score : 0.48613 / Kaggle rank : 1,357 / 3,251 (41.7%)
Github : https://github.com/kish191919/Bike-Sharing-Demand_by_R