

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](https://www.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

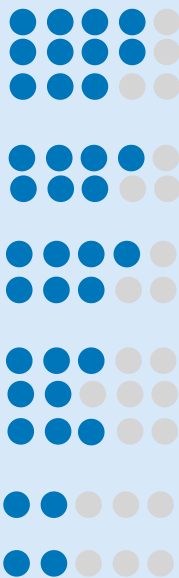
Hive

Framework

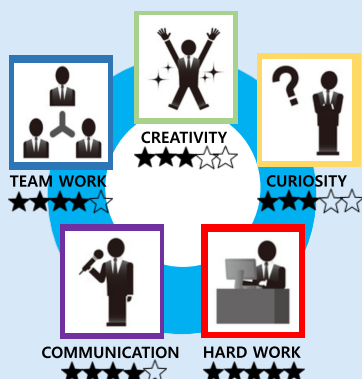
Flask

Cloud

AWS



PERSONAL



EDUCATION

- **FAST CAMPUS** - Seoul, Korea
Data Science Academy

JAN 2018 - MAY 2018

- **UDACITY** - Online Course
Intro to Programming Nano degree

JUL 2017 - AUG 2017

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

MAR 2000 - FEB 2009



PROJECT

Portfolio Website: <http://datavoyagerdanny.com>

- **Predict Used-Car Price in Georgia** [Service Website : <http://dannyki.ga/>]
Crawled on cars.com to collect data and store it in AWS's mysql.
After preprocessing the stored data, it is learned by putting it in the machine learning model to predict the used car price.
The models learned on the AWS server are implemented as web services using the Flask web framework.
- **[KAGGLE Competition] Predict House Prices / Regression**
Developed and applied OLS algorithms to predict house price in Ames, Iowa.
It was the first project I submitted to Kaggle Competition and solved the problem with a probabilistic approach.
- **[KAGGLE Competition] Spooky Author Identification / Text Classification**
As a text analysis project, it is a problem to see which author is writing the article. After vectorizing the words, classify them by machine learning with Naive Bayes Classification.
- **[KAGGLE Competition] Titanic Machine Learning from Disaster / Classification**
Predicted survival on the Titanic.
- **[KAGGLE Competition] Bike Sharing Demand / Regression**
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 Assistant Manager

AUG 2015 - DEC 2017

Dongwon Autopart Technology Georgia LLC (web site)

Hogansville, Georgia

Establish the negotiating strategy for sales and took the lead position in markets

Purchasing Assistant Manager

DEC 2013 - JUL 2015

Kukdo Chemical Co.,LTD. (web site)

Seoul, Korea

Analyze price proposals, financial reports, market and other data

Purchasing Specialist

OCT 2011 - NOV 2013

Lotte Chemical Alabama Corp (web site)

Auburn, Alabama

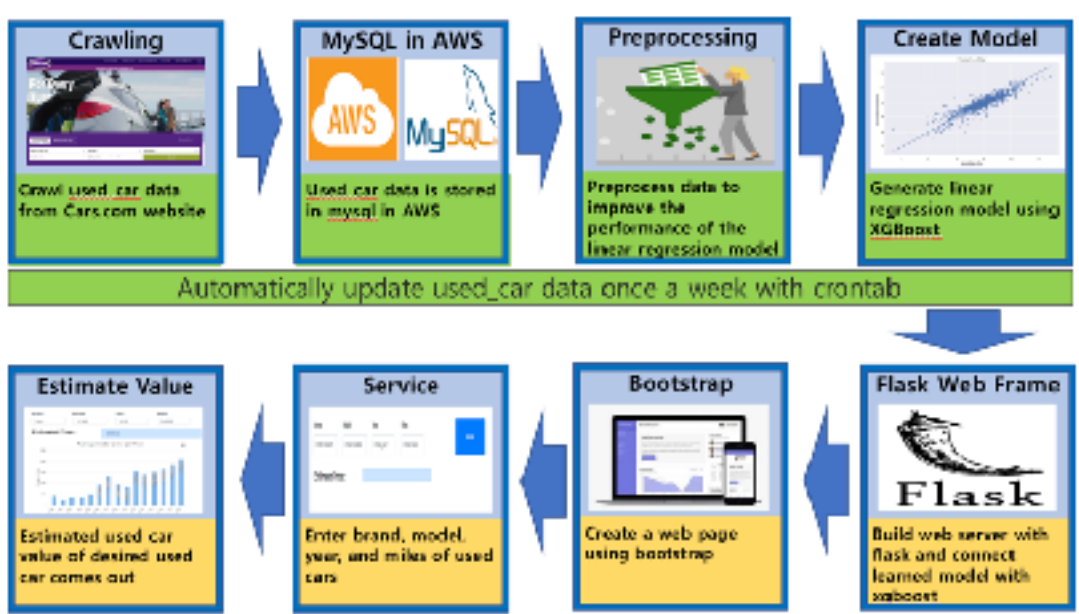
Plan and purchase chemical materials from domestic and imported with economic conditions



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, Iowa
 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						
[Note: coefficients are in millions]						
Dep. Variable:	SalePrice	R-squared:	0.948			
Model:	OLS	Adj. R-squared:	0.947			
Method:	Least Squares	Root Mean Square:	403.4			
Date:	Mon, 26 Mar 2019	Prob (F-statistic):	0.00			
Time:	01:13:10	Cup-Sizes:	1405.0			
No. Observations:	1460	ADF:	-2.098			
df Residuals:	1326	BIC:	-2328.			
df Model:	16					
Variance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	12.1818	0.355	34.295	0.000	11.476	12.888
1[Neighborhood][T.McMans]	-0.3265	0.369	-0.882	0.378	-0.163	0.110
2[Neighborhood][T.McMans]	-0.3495	0.377	-0.927	0.358	-0.133	0.133
3[Neighborhood][T.McMans]	-0.3818	0.371	-1.030	0.308	-0.169	0.156
4[Neighborhood][T.McMans]	-0.3424	0.372	-0.923	0.357	-0.130	0.139

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 110 276]
 [ 531 4764 340]
 [ 100  0 5347]]
```

10-fold Cross Validation Report:				
	precision	recall	f1-score	support
0	0.86	0.94	0.90	7980
1	0.96	0.85	0.90	5635
2	0.99	0.89	0.94	5944
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



PROJECT

(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 Classification, VotingClassifier
Model : **VotingClassifier Model**

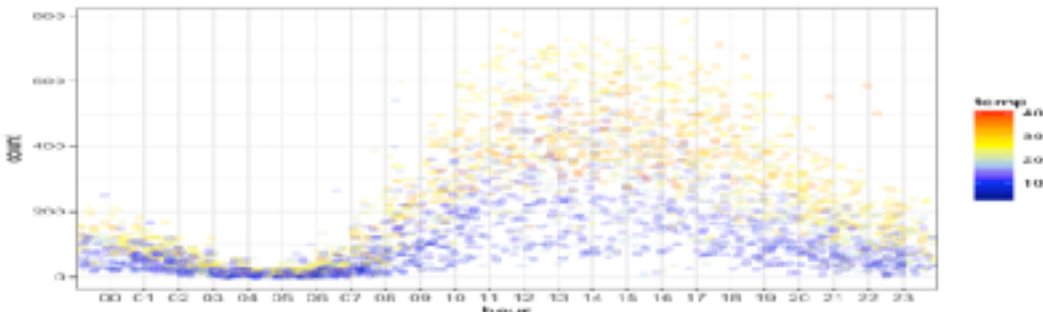
```
Confusion Matrix :  
[[484  57]  
 [ 60 260]]  
  
10-fold Cross Validation Report:  
      precision    recall  f1-score   support  
  
     0       0.86      0.85      0.86      541  
     1       0.82      0.76      0.79      340  
  
avg / total       0.84      0.84      0.84      881
```

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