MyungHoon Jin

A.I researcher

Address 183-19, Munsi-ro, Osan-si, Gyeonggi-do, Republic of Korea

Phone +82) 4011-4990

GitHub https://github.com/jinmang2

Email jinmang2@gmail.com

Profile

Rubato Lab에 지원하게 된 `진명훈` 이라고 합니다. 가천대에서 금융수학을 전공했으며 18년도부터 Data Science, AI에 관심을 가지고 공부했고 현재는 Persona AI에서 딥러닝 연구개발 사원으로 근무 중 입니다. 주로 Seq2Seq 기반의 자연어 모델을 사용해봤고 Sequential 한 data들에 대한 연구를 하고 싶은 꿈이 있습니다. 음성 합성 또한 제가 공부하고 다뤄보고 싶은 연구였고 이번 Lib project 기회가 생겨 지원했습니다. 제 이력 및 세부 사항은 아래에 기술하도록 하겠습니다.

Research Experience

Sep 2019 - present
남부터미널

Persona A.I. R&D researcher

챗봇 딥러닝 연구개발 사원

- 한화손해보험 FAQ 데이터 활용 Seq2Seq 대화모델 개발 (中)
- Seq2Seq 관련 논문 리뷰
- Joint Sentiment Topic Modeling을 활용한 다중 감정 분류기 개발
- https://github.com/jinmang2/t2snet

Sep 2017 – May 2019
성남, 가천대

Gachon Convergence Research Center

경영대학 장유상 교수 연구실 조교 (Panel, Sequential Data Handling)

- 국가 대도시별 미세먼지(PM2.5) 데이터를 Experience Curve 기법으로 분석
- 세계 국가별 환경 데이터 catch-up 현상 분석을 위한 convergence analysis 실시
- Panel data 분석을 위해 Stata PCSE 모델을 활용, 미세먼지 데이터 분석
- 논문 출판 `Convergence Analysis of the Sanitation Index for 158 Countries`

Sep 2018 – May 2019
성남, 가천대

Sequential Analysis on stock market data

금융수학학부 문경숙 교수 머신러닝 프로젝트 참여

- Tree 기반 ensemble 모형 XGBoost, LightGBM을 활용
- 주식 수익률 최대화를 objective로 설정
- 이를 위해 다양한 target variable functio을 실험

$$Y_{t,n} = Step\left(\ln\left(\frac{Close^{s}_{i+t}}{\frac{1}{n}\sum_{k=0}^{n-1} Close^{s}_{i-k}}\right)\right) \quad for \begin{cases} t \in range(1,31,3) \\ n \in range(1,89,3) \end{cases}$$

“Sequential Data Analyst”

Paper Reviews

https://github.com/jinmang2/Paper-Review

- Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation
  - Seq2Seq, GRU 유닛 소개
- Neural Machine Translation by Jointly Learning to Align and Translate
  - Bahdanau Attention 제안
- Effective Approaches to Attention-based Neural Machine Translation
  - Luong Attention 제안
- A Structured Self-Attentive Sentence Embedding
  - Self-Attention 제안 (vector representation -> matrix representation)
- Attention Is All You Needs
  - Transformer, first transduction model with self-attention without recurrent and convolution
- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
  - Transformer Encoder, Transfer Learning, MLM and NSP, fine-tune based (not a feature-based)
- Deep contextualized word representations
  - Feature-based transfer learning, ELMo. BiLM을 사용하지만 이를 독립적으로 concat
- Highway Networks
  - Residual Connection의 상위 버전. LSTM에 영감을 받음
- Unsupervised Pre-training of a Deep LSTM-based Stacked Autoencoder for Multivariate Time Series Forecasting Problems
  - 시계열 예측을 위해 LSTM Stacked-AE 모형 제안
- Learning Spatiotemporal Features with 3D Convolutional Networks
  - 3D Convolution pre-training 모델
- Real-World Anomaly Detection in Surveillance Videos
  - 위 C3D를 활용하여 감시카메라 이상탐지 실시. Ranking loss 및 MIL 기법 활용
- Neural Machine Translation of Rare Words with Subword Units
  - BPE를 NMT에 적용! 새로운 버전으로 BPE 알고리즘을 제시
- A Neural Probabilistic Language Model
  - Bengio 교수가 제안한 neural network 기반 language model

"자연어 위주로 리뷰했지만"

"음성 데이터 또한 관심을 가지고 있고 추가로 논문 리뷰할 계획입니다."

Education

Mar 2013 – Feb 2019

Gachon University

- Major: Financial Mathematics (GPA: 3.93 / 4.5)

Jun 2019 – Aug 2019

Korean Standards Association

- Machine Learning & Deep Learning, National support

Study History

Mar 2014 – Dec 2014

Linear Algebra

- 학부 전공으로 공부
- 교재: Elementary Linear Algebra 10th Edition

Mar 2018 – Jun 2018

Python

- Mooc python 강의 – 가천대 산업경영공학과 최성철 교수

Mar 2018 – Aug 2018

Machine Learning

- 데이터 융합 전공 청강
- 교재: Hands on Machine Learning
- 최성철 교수 Machine Learning 강의

Feb 2020 – present

Korean Embedding

- 자연어 처리 스터디
- 논문 리뷰 및 구현
- 교재: 이기창님 한국어 임베딩
- https://github.com/jinmang2/KoreanEmbedding

Apr 2020 – present

Optimization

- 선형대수 복습 및 딥러닝 수학적 베이스를 키우기 위해 스터디
- 교재: Numerical Optimization
- https://github.com/jinmang2/optimization

"Sequential Data를 주로 다루는 수학적 Base가 탄탄한"

"AI Researcher가 되기 위해"

"위와 같이 공부하며 시간을 보냈습니다."

Projects

- NCSOFT User Churn Prediction
  - Big Contest 2019 Champions League, 1차 통과
- Classify Default Credit Card
  - UCI 데이터를 활용, 신용 파산 예측 문제 해결
- Lion/Tiger/Leopard Classifier
  - 동물 이미지를 crawling하여 CNN으로 분류 모델 구축
- Stock recommend algorithm with precision
  - KOSPI 과거 주식 데이터를 기반으로 portfolio 제안 모델 구축
- Airbnb New User Booking
  - Kaggle의 Airbnb 다중 분류 문제를 Tree기반 모델로 해결
- Catch-up effect on EPI2016 Data Set
- Option Pricing & Delta Hedging Portfolio
- Portfolio Optimization with CAPM
- Hedge Strategy by Derivative (Futures)

"제가 학부 시절부터 한 프로젝트를 정리한 내용입니다."

"Rubato Lab의 task와 직접적인 연관은 없지만"

"제가 어떤 활동을 해왔고 지금까지의 족자를 말씀드리고자 합니다."

Projects #1

AUG 2019-present

Team project (3)

#1 NCSOFT User Churn Prediction

GitHub: github.com/jinmang2/ncsoft_predict_churn

Summary

- Champions League of Big Contest 2019
- *Data*: activity, pledge, trade, combat, payment data by characters of each user for 28 days
- *Define Churn*: users whose connection history has been lost for more than 7 consecutive days during 64 observation days
- To maximize ncsoft's expected returns, forecast the followings;
  - ✓ $\hat{T} = predicted\ survival\ time, 1{\sim}63{:}\ churn\ |\ 64{:}\ remain$ (fig1-1)
  - ✓ $\hat{R} = predicted\ daily\ amount\ spent, positive\ real\ number$ (fig 1-2)
- *Objective function*: $\hat{E}(r) = \gamma \times \hat{T} \times \hat{R} - C, \gamma{:}\ conversion\ rate, C{:}\ cost$
- Adjust 3 ideas as followings;
  - ✓ Flatten the weekly variable to make it a features (fig 1-3)
  - ✓ Calculate the count variable for each features to become robust over time (fig 1-4)
  - ✓ Change labels to form; $yln(y)$ (fig 1-5)
- Deriving the top 30 on the Leader Board with 14,000 points
- 1st pass on Big Contest 2019

Role within the Team

- Build baseline by flatten data set
- Create unity and joyful atmosphere as a younger brother
- Providing ideas, change labels

I felt these things

- Not only sequential model, but also cross-sectional model is important
- Importance of target distributions
- Effect of team synergy
- Reducing user churn is important in business problems because it increases expected returns

Figure 1-1. User Churn ratio and frequency by survival time

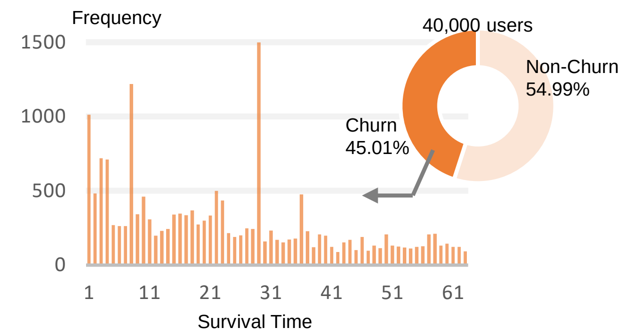


Figure 1-2. amount spent by survival time

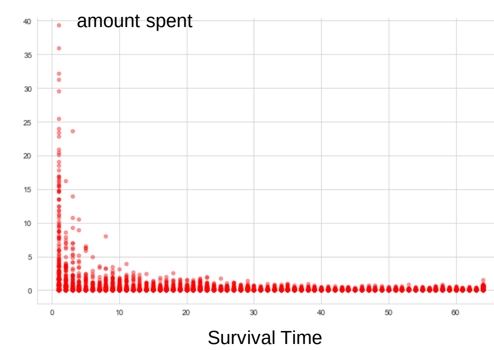


Figure 1-3. Flatten the weekly variable to make it a feature

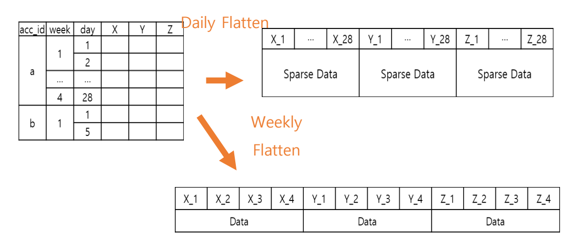


Figure 1-4. Calculate count variable for each features

acc_id	day	X	Y
a	1	0.0011	0.0023
	2	NaN or 0	0.4591
	3	0.0320	0.3336
	5	0.0220	NaN or 0
	7	0.1400	NaN or 0
	9	1.2780	0.4802
	16	NaN or 0	0.6911
	21	0.9000	NaN or 0
	26	0.5000	NaN or 0
	28	0.2760	NaN or 0

acc_id	Avg(X)	Avg(Y)	Count(X)	Count(Y)
a	0.3936	0.3932	8 / 10	5 / 10

Figure 1-5. Equations of changing label

Survival Time

$$\hat{T} = \begin{cases} (\hat{T} - 32) \times \ln(32 - \hat{T}), & if\ \hat{T} < 32 \\ (\hat{T} - 32) \times \ln(\hat{T} - 31), & if\ \hat{T} \geq 32 \end{cases}$$

$$\hat{T} = \begin{cases} 64, & if\ \hat{T} \geq 64 \\ 1, & if\ \hat{T} \leq 1 \\ \hat{T}, & otherwise \end{cases}$$

Amount Spent

$$\hat{R} = \hat{R} \times \ln(\hat{R} + 1)$$

$$\hat{R} = \begin{cases} \hat{R}, & if\ \hat{R} \geq 0 \\ 0, & otherwise \end{cases}$$

Projects #2, #3

JUL 2019-
JUL 2019

Side project
(1)

#2 Classify Default Credit Card (UCI Data Set)

GitHub: github.com/jinmang2/KSA_Modules/tree/master/perform_eval/2nd

Summary

- KSA 2nd performance test
- *Data*: personal information(gender, education,, age, etc.) and past 6 months consumption and default history
- *Class imbalance problem:* Trying SMOTE and stratified sampling (fig 2-1)
- Propose XGB(82.29%), CNN(81.57%), Voting Classifier(81.91%)

Do as Followings

- Do EDA for get feature vectors
- Perform test; SMOTE vs stratified sampling
- Feature selection with feature importance gained by XGB (fig 2-2)
- Solve classification problem with CNN by keras

I felt these things

- Use CNN, solve classification problem on cross-sectional data (fig 2-3)
- In order to solve the imbalance class problem, a strategy should be devised rather than simply applying SMOTE.

Figure 2-1. Ratio of Binary Target Variable

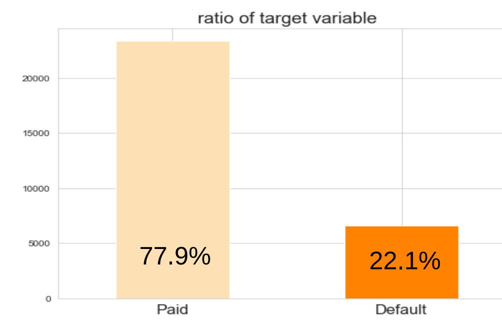


Figure 2-2. Feature Importance of XGB Classifier

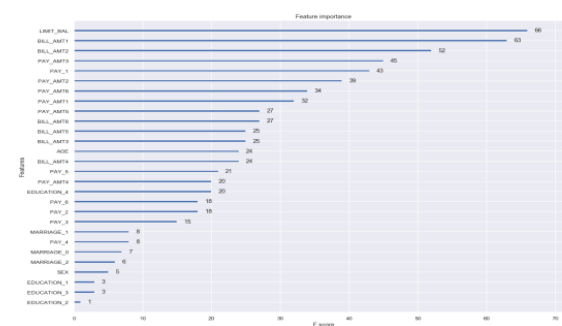
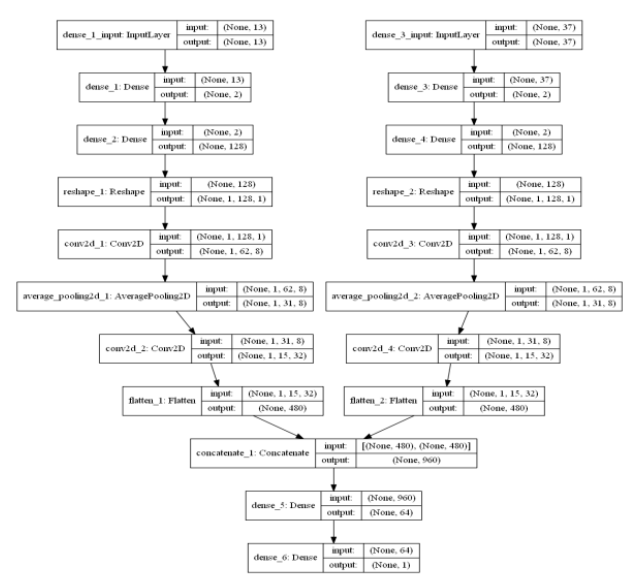


Figure 2-3. CNN Structure by python-graphviz



JUL 2019-
JUL 2019

Team project
(3)

#3 Lion/Tiger/Leopard/etc. Classification

GitHub: github.com/jinmang2/animal_classifier

Summary

- CNN project , KSA module 6
- *Data*: Gather lion, tiger, and leopard image from google by web crawling
- Train 1,500, valid 450, test 3,773 Image (300*300*3)
- Only 10% Images are directly labeled and the rest of the labeling is automatically done with binary classification CNN
- And then, quadruple classification of labeled data (fig 3-1) (Lion, Tiger, Leopard, etc.)
- Propose test accuracy 85%, recall 88%

Role within the Team

- Conduct and plan roles for each team member as a leader
- Leopard data collection and binary classification
- Run and test final model

I felt these things

- As we gathered the data, we saw why the data collection and preprocessing took so long in machine learning projects.
- Visualize feature maps to see patterns for each animals (fig 3-2)

Figure 3-1. CNN Structure on Binary & Quadruple model

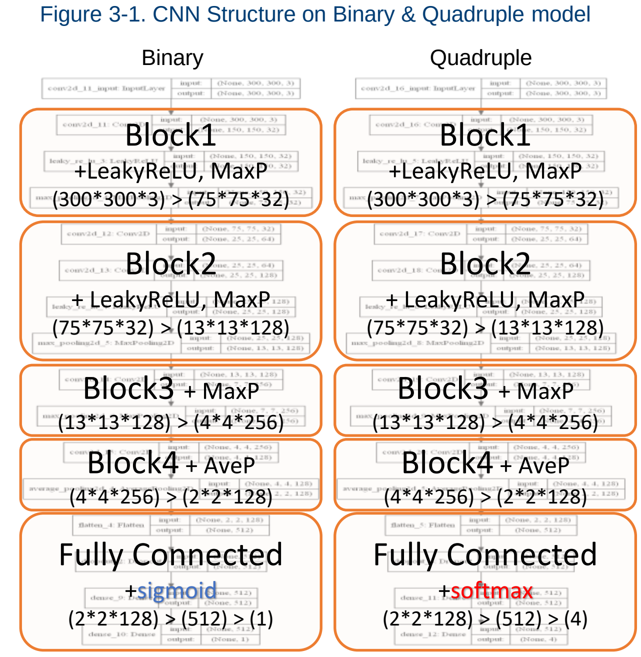
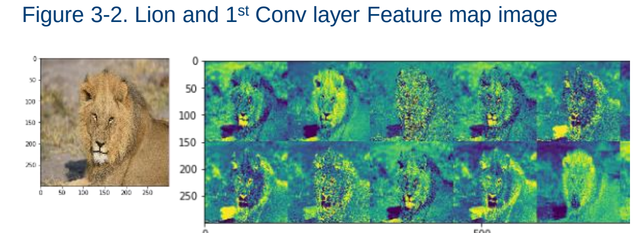


Figure 3-2. Lion and 1st Conv layer Feature map image



Projects #4, #5

JUN 2018-
MAY 2019

Team project
(2)

#4 Stock recommend algorithm with precision

GitHub: github.com/jinmang2/stock_recommender

Summary

- Introduce stock recommend algorithm with precision
- *Data*: Top 16 stock by market category on KOSPI, '07.01.01~'18/06/30
- Price and technical indicator features such as MA, MACD, RSI, etc. were used.
- Set 300 target variable, (t: predict period, n: time window)
  - $Y_{t,n} = Step\left(\ln\left(\frac{Close_{i+t}^{s}}{\frac{1}{n}\sum_{k=0}^{n-1} Close_{i-k}^{s}}\right)\right)\ for \begin{cases} t \in range(1,31,3) \\ n \in range(1,89,3) \end{cases}$
- Do these following steps;

✓ Calculates the time window of the technical indicators with the highest correlation for each target variable

✓ Select t^* which has best performance on 70% train data

✓ For each $n_1, n_4 \sim n_{88}$, the right to vote is forfeited if the precision of 90 days is smaller than 0.7 and the $\hat{Y}_{t^*,n^*}$ is obtained by holding a vote.

✓ If $\hat{Y}_{t^*,n^*} = 1$, buy stocks with $\frac{1}{N}$ shares (N is number of stocks which rise)

✓ If not, unwinding position. (result; fig(4-1))

Role within the Team

- Calculate training features and target variables
- Test step 2~4 and modularize written code

I felt these things

- Importance of sequence and time shift in time series data
- Pricing data is not enough to predict stocks

Figure 4-1. Returns and Accuracy for 16 stocks

표 3. 16개 주식에 대한 수익률과 평가지표

	수익률	보유 수익률	정밀도	정확도
KB 금융	-0.48%	-8.87%	33.33%	56.85%
LG	-4.39%	-6.96%	88.57%	56.43%
LG 생활건강	23.37%	33.96%	84.62%	45.64%
LG 화학	44.45%	13.63%	70.00%	68.46%
NAVER	4.30%	-9.38%	28.00%	52.28%
SK	-3.29%	-7.08%	28.57%	51.04%
SK 이노베이션	21.66%	24.25%	85.71%	65.56%
SK 텔레콤	5.17%	-13.25%	60.47%	56.85%
SK 하이닉스	27.77%	24.02%	54.12%	48.55%
삼성생명	20.42%	-17.31%	47.73%	68.88%
삼성전자	18.70%	-1.89%	61.25%	63.45%
삼성중공업	26.33%	-44.27%	61.76%	66.39%
셀트리온	109.21%	98.94%	79.05%	64.32%
신한지주	11.24%	-12.98%	70.37%	69.71%
현대건설	51.65%	22.21%	25.64%	48.96%
현대글로비스	18.58%	-30.70%	56.25%	63.49%

APR 2018-
JUN 2018

Team project
(3)

#5 Airbnb New User Bookings (Kaggle)

GitHub: github.com/jinmang2/airbnb_new_user_bookings

Summary

- Multi-class problem that predicts which country the first user will travel based on data provided by Airbnb (fig 5-1)
- The random forest was used to provide 83% accuracy in predicting tests. (fig 5-2)

Role within the Team

- Tree-based ensemble model Hyper parameter tuning
- Responsible for introduction and model building at the final PT

I felt these things

- I have studied various kernels of Kaggle.
- As my first machine learning project, I built a baseline and went through the overall process.

Figure 5-1. Training Destination Distribution by Gender

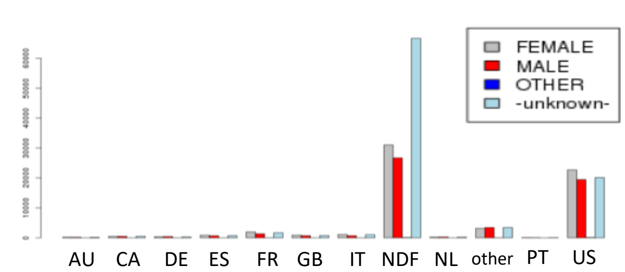


Figure 5-2. Confusion Matrix of Final Model

predict / actual	AU	CA	DE	ES	FR	GB	IT	NDF	NL	PT	US	other
AU	488	0	0	0	2	0	0	41	0	0	8	0
CA	0	1211	0	2	4	1	1	176	1	1	27	4
DE	0	1	933	0	0	0	2	98	0	0	25	2
ES	0	1	0	1926	1	0	1	266	0	0	51	3
FR	1	1	0	5	4294	1	2	590	3	0	116	10
GB	0	2	1	0	0	1972	0	287	0	0	56	6
IT	0	1	0	2	2	4	2347	395	1	0	75	8
NDF	7	17	6	28	68	33	47	122574	10	3	1579	171
NL	0	0	0	0	2	0	0	88	653	0	16	3
PT	0	0	0	0	0	2	0	23	0	186	5	1
US	6	7	9	24	43	24	27	6303	8	2	55816	107
other	1	2	3	6	15	2	7	1353	3	0	274	8428

Projects #6, #7

SEP 2017-
DEC 2017

#6 Catch-up effect on EPI2016 Data Set

GitHub: github.com/jinmang2/gachon_research

Summary

Side project
(1)

- Projects conducted by Gachon Convergence Research Center
- Studying whether developing countries catch up with developed countries on the environmental indexes provided by Environmental Performance Index 2016. (fig 6-1)
- Do convergence analysis with sigma ang gamma Indexes (fig 6-2, 6-3)

Do as Followings

- Subgroup analysis by income and region
- Create figure and table for article
- T-test and chi-squared statistical tests.

I felt these things

- Not only big-data but also small data on data analysis
- Importance of organize results into table and presentation

Figure 6-1. Averaged Sanitation Indexes

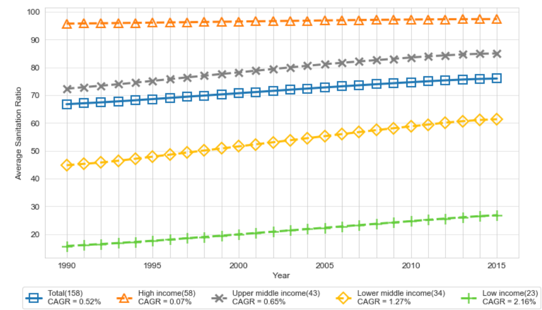


Figure 6-2. Normalized Sigma Indexes on Sanitation

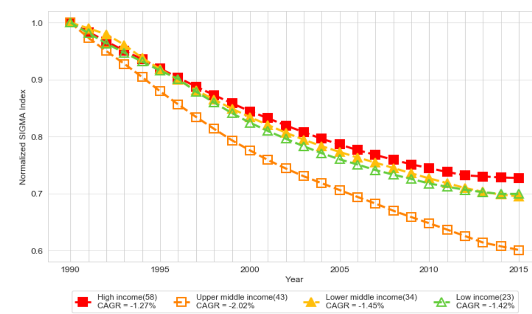
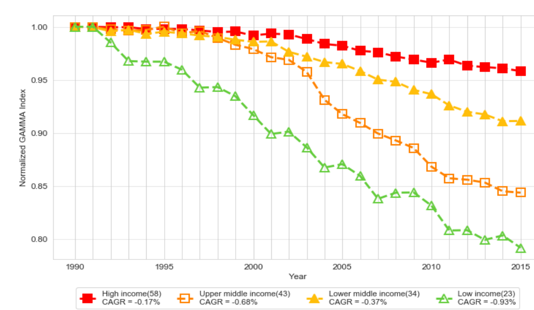


Figure 6-3. Normalized Gamma Indexes on Sanitation



OCT 2017-
DEC 2017

#7 Option Pricing & Delta Hedging Portfolio

GitHub: github.com/jinmang2/option_valuation

Summary

Team project
(2)

- Build a delta hedging portfolio by KOSPI200 (Nov 17) call / put index option & futures and analyzes P&L with Greeks
- Since we predicted a low volatility market, we configured our portfolio as follows: (fig 7-1)
  - ✓ Short C330, C340, P322.5, P330; 676:1692:676:676 contracts
  - ✓ Long C335, C345; 1692:676 contracts
- Since gamma<0, portfolio is showed overall short gamma position.
- In addition, long positions were taken to compensate for fluctuations in the market.
- But geopolitical risks, such as washing machine tubes raising base rate, have maximized market volatility and loss on 11/27, 11/30, 12/4, 12/6.
- Fortunately, nothing happened on Quadruple Witching Day, so we benefited from theta's time value, totaling ₩ 11,125,000 (fig 7-2)

Role within the Team

- Calculate daily P&L and Greeks change
- Analyze causes of portfolio's P&L by Greeks (Use taylor expansion)

I felt these things

- Impact of geopolitical risks on portfolio
- Importance of unit. For instance, vega and theta (adjust business day)
- P&L analysis by applying Black-Scholes formula

Figure 7-1. Pay-off graph for our option portfolio

Table 7-1. Cumulative P&L our portfolio

Date	Daily P&L	Cumulative P&L	Greeks P&L
2017.11.10	₩ -	₩ -	₩ -
2017.11.13	-₩ 14,765,000	-₩ 14,765,000	-₩ 34,608,194
2017.11.14	₩ 81,245,000	₩ 66,480,000	₩ 26,019,513
2017.11.15	-₩ 80,402,500	-₩ 13,922,500	-₩ 31,046,395
2017.11.16	₩ 70,465,000	₩ 56,542,500	₩ 67,685,166
2017.11.17	₩ 15,345,000	₩ 71,887,500	₩ 16,564,104
2017.11.20	₩ 5,055,000	₩ 76,942,500	-₩ 24,045,784
2017.11.21	₩ 38,090,000	₩ 115,032,500	₩ 33,704,130
2017.11.22	₩ 1,260,000	₩ 116,292,500	-₩ 34,860
2017.11.23	-₩ 9,725,000	₩ 106,567,500	₩ 6,713,021
2017.11.24	₩ 1,490,000	₩ 108,057,500	₩ 19,131,648
2017.11.27	-₩ 161,100,000	-₩ 53,042,500	-₩ 153,700,556
2017.11.28	₩ 23,350,000	-₩ 29,692,500	₩ 39,512,040
2017.11.29	₩ 9,290,000	-₩ 20,402,500	-₩ 972,604
2017.11.30	-₩ 228,332,500	-₩ 247,372,500	-₩ 212,058,608
2017.12.01	₩ 73,240,000	-₩ 174,132,500	₩ 36,710,446
2017.12.04	-₩ 72,145,000	-₩ 246,277,500	-₩ 91,702,825
2017.12.05	₩ 17,240,000	-₩ 229,037,500	₩ 35,155,877
2017.12.06	-₩ 151,307,500	-₩ 380,345,000	-₩ 193,376,626
2017.12.07	₩ 44,430,000	-₩ 335,915,000	₩ 44,407,402
2017.12.08	₩ 90,250,000	-₩ 245,665,000	₩ 83,326,125
2017.12.11	₩ 72,375,000	-₩ 173,290,000	₩ 73,917,259
2017.12.12	₩ 57,725,000	-₩ 115,565,000	₩ 37,645,121
2017.12.13	₩ 45,220,000	-₩ 70,345,000	-₩ 6,910,066
2017.12.14	₩ 81,470,000	₩ 11,125,000	₩ 13,545,040

Projects #8, #9

OCT 2017-DEC 2017

Team project (4)

#8 Portfolio Optimization with CAPM

GitHub: github.com/jinmang2/portfolio_optimization

Summary

- Use Markowitz's portfolio optimization theory and CAPM to build a portfolio that maximizes CAL slope as a KOSPI stock and provide quantitative and qualitative reasons
- Proposal Portfolio: Foosung (093370), Asiana airline (020560), S-Oil Corp (010950) (fig 8-1, fig 8-2)
- Expected returns: 20.7% / yr
- Operating income: 3.5% / 22 Days = 40.15% / yr (fig 8-3)

Role within the Team

- build an overall process as a leader
- Validate number of various cases with excel to find $\text{argmax}_s\, CAL$

I felt these things

- I was fascinated by the way the portfolio was optimized through statistical methods such as standard deviation, expected value, etc.
- Felt limited in finding $\text{argmax}_s\, CAL$ with excel and the needs to learn other programming languages

OCT 2017-DEC 2017

Team project (4)

#9 Hedge Strategy by Derivative (Futures)

GitHub: github.com/jinmang2/portfolio_optimization

Summary

- Set a strategy to maximize return on risk by composing a portfolio of stocks that have a negative correlation with foreign commodity futures
- Proposal portfolio (take long position on futures): (fig 9-1)
    1. 25%; Carbon Emission Futures (CFI2Z9) in ICE
    2. 40%; Yanzhou Coal Mining Co Ltd (600188, SH)
    3. 35%; Vestas Wind Systems A/S (VWS)
- Carbon emission is negative correlated with two stocks.
- Propose expected return: 6.0% / 1yr
- Team project 1st on 6 teams

Role within the Team

- Propose carbon emission futures and investigate price variables
- Suggest qualitative content to explain quantitative figures of selected portfolio and test case study. (fig 9-2)
- Investigate relation of carbon emission, fossil and renewable energy

I felt these things

- Improving the quality of the content and presenting it to the client
- See how mathematics is used in risk management

Figure 8-1. Optima Portfolio

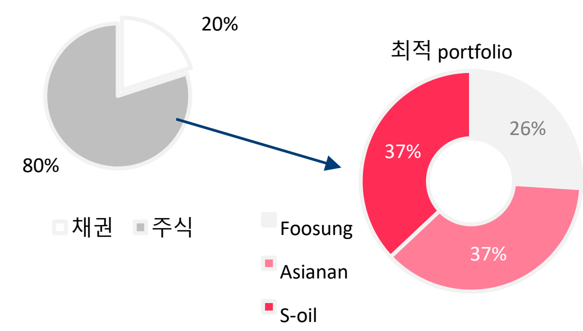


Figure 8-2. Efficient Frontier of 3 stocks and 6 weights point

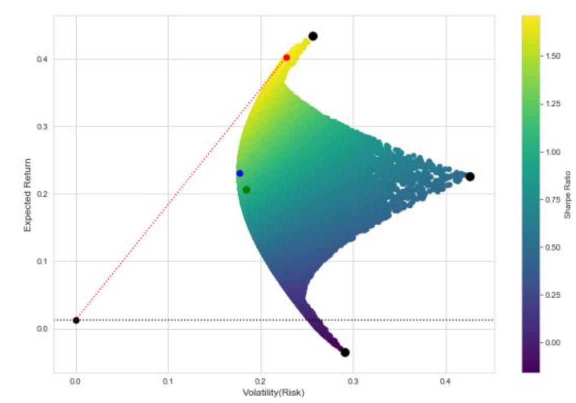


Figure 8-3. Cumulative Returns for 6 portfolio weights

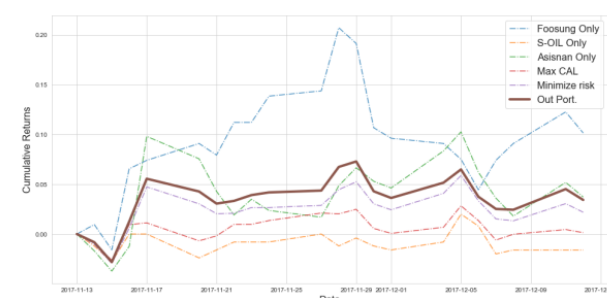


Figure 9-1. Bar Chart of portfolio weight ratio

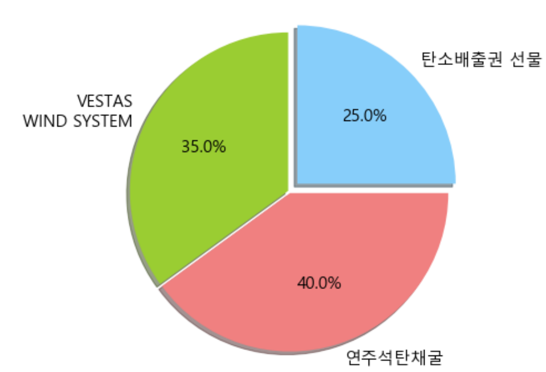


Figure 9-2. Analyze Case study of tail risk

브렉시트(2016/6/23~7/22)		
베스타스	연주석탄	탄소
-1%	16%	-14%
헤지된 수익률		
3%		
베스타스 급락시(2016/9/18~12/8)		
베스타스	연주석탄	탄소
-24%	-5%	40%
헤지된 수익률		
-1%		
연주 급락시 (2015/8/19-8/24)		
베스타스	연주석탄	탄소
-16%	-20%	3%
헤지된 수익률		
-13%		

Appendix

Figure 1-1. User Churn ratio and frequency by survival time

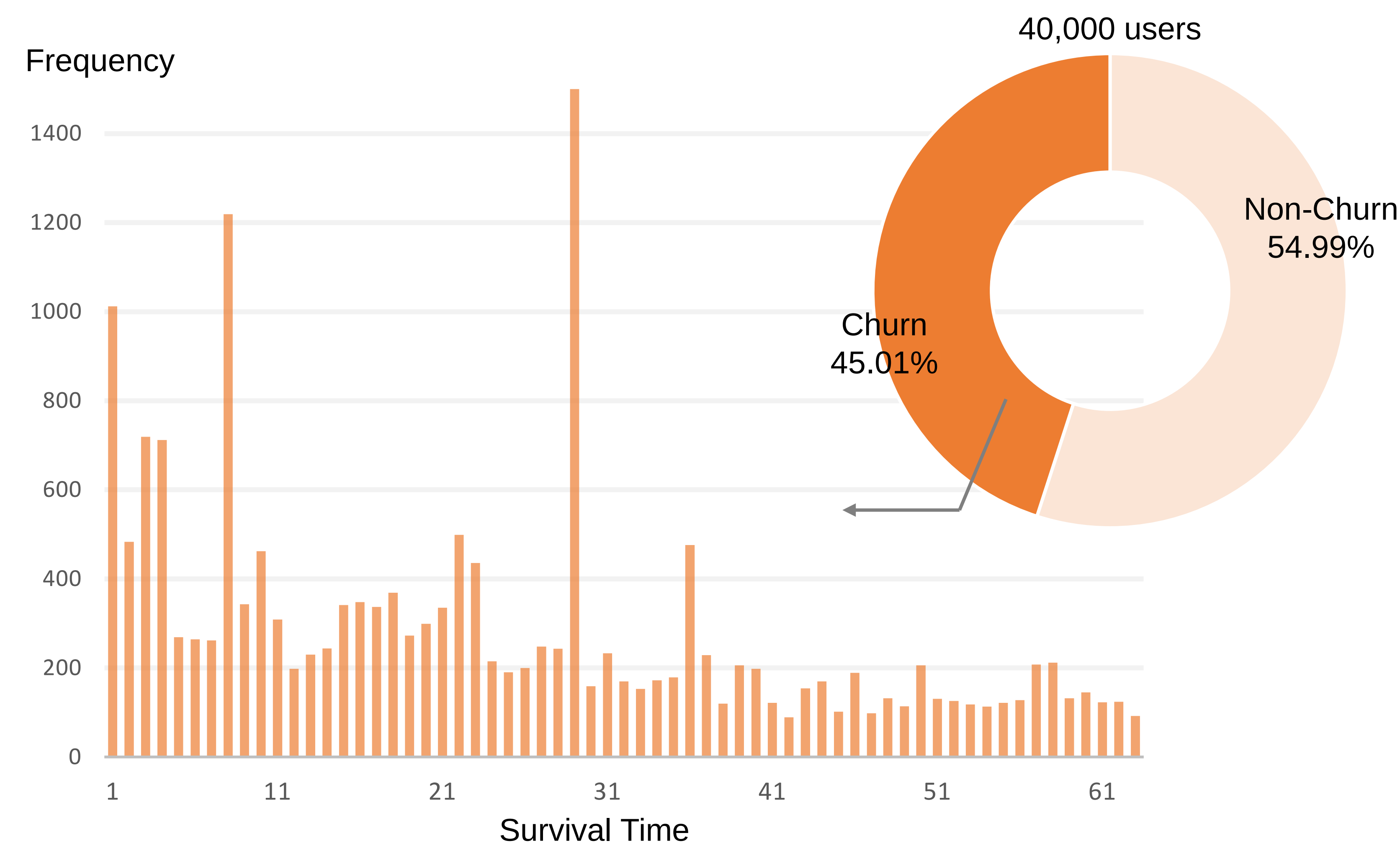
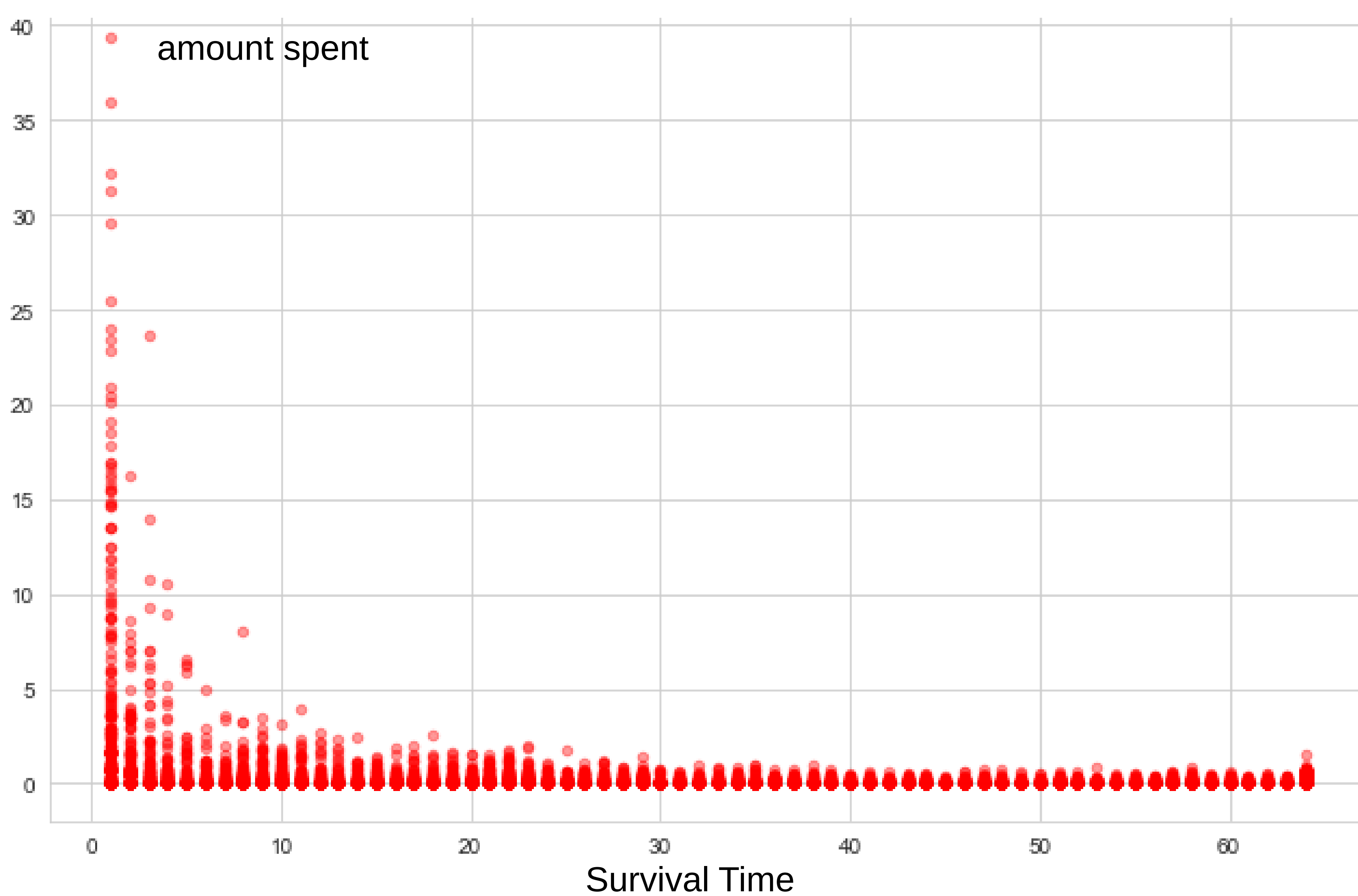


Figure 1-2. amount spent by survival time



Appendix

Figure 1-3. Flatten the weekly variable to make it a feature

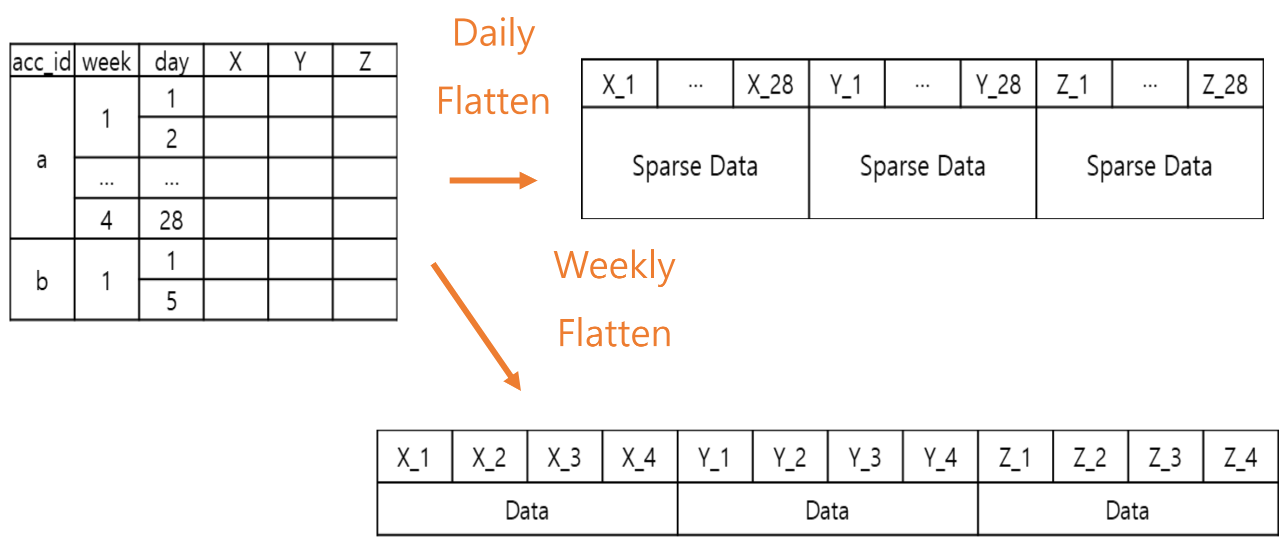


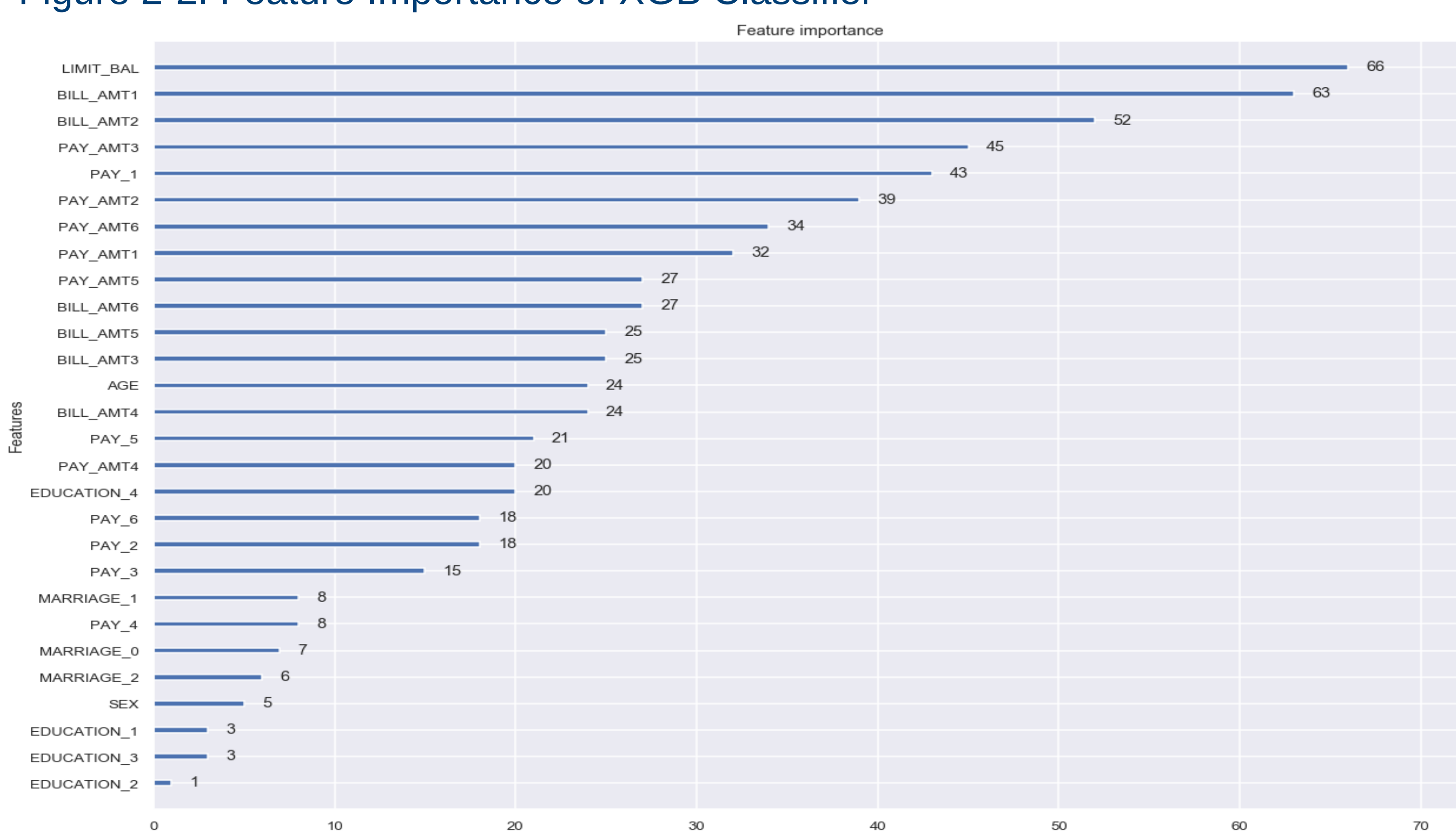
Figure 1-4. Calculate count variable for each features

acc_id	day	X	Y
a	1	0.0011	0.0023
	2	NaN or 0	0.4591
	3	0.0320	0.3336
	5	0.0220	NaN or 0
	7	0.1400	NaN or 0
	9	1.2780	0.4802
	16	NaN or 0	0.6911
	21	0.9000	NaN or 0
	26	0.5000	NaN or 0
	28	0.2760	NaN or 0

→

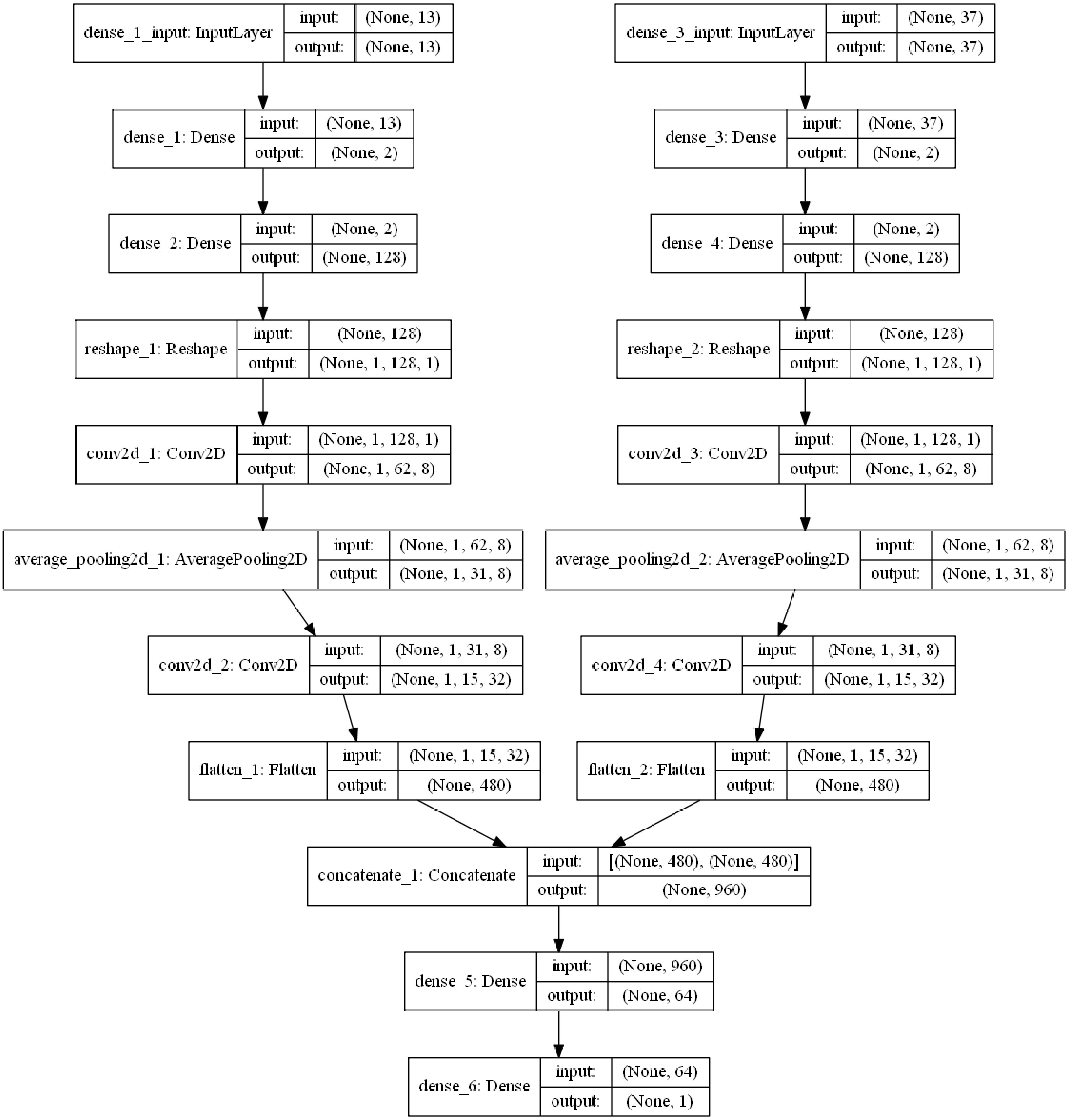
acc_id	Avg(X)	Avg(Y)	Count(X)	Count(Y)
a	0.3936	0.3932	8 / 10	5 / 10

Figure 2-2. Feature Importance of XGB Classifier



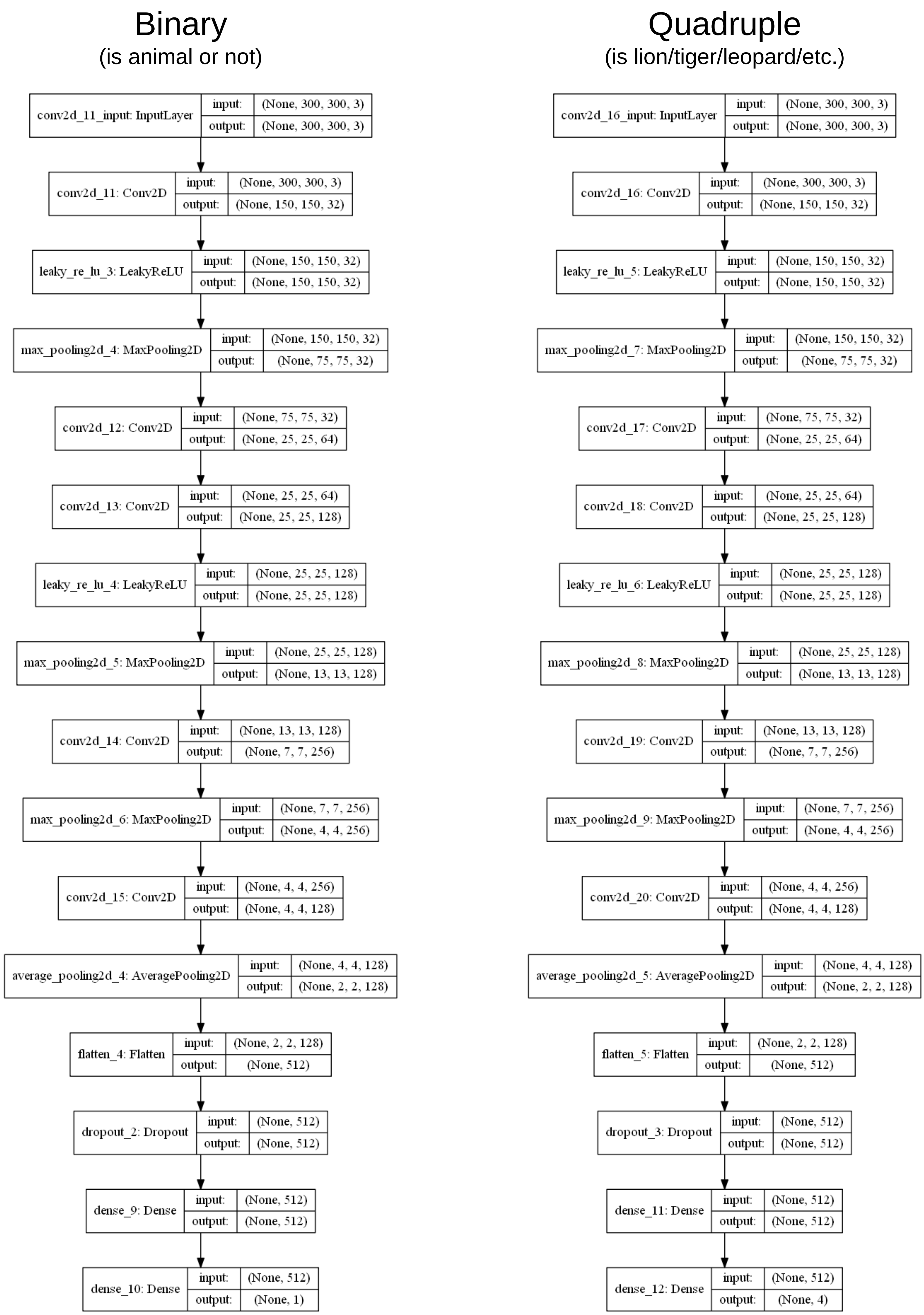
Appendix

Figure 2-3. CNN Structure by python-graphviz



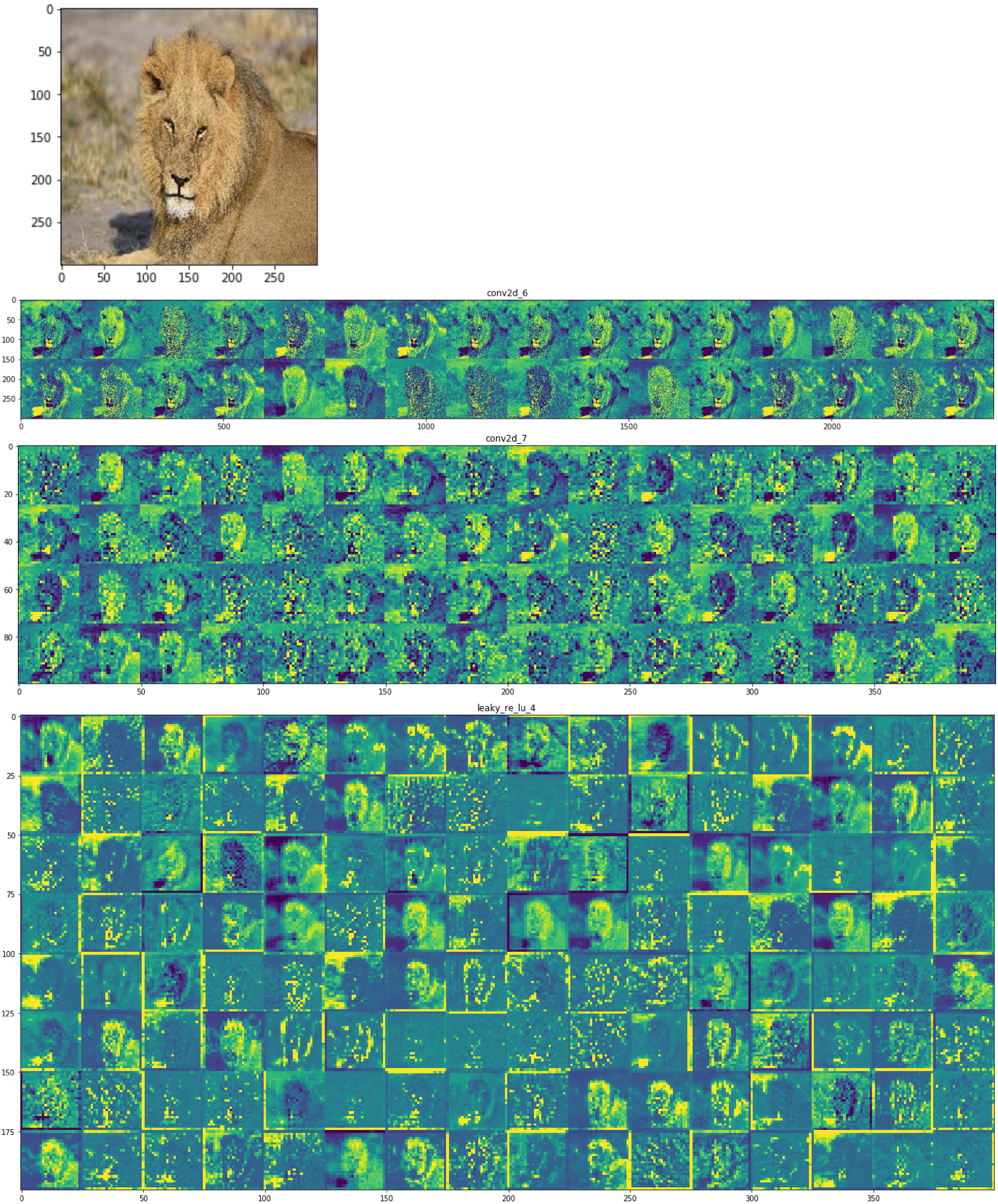
Appendix

Figure 3-1. CNN Structure on Binary & Quadruple model



Appendix

Figure 3-2. Origin Lion and 1st , 4th , 6th Conv layer Feature map image



Appendix

Figure 5-1. Training Destination Distribution by Gender

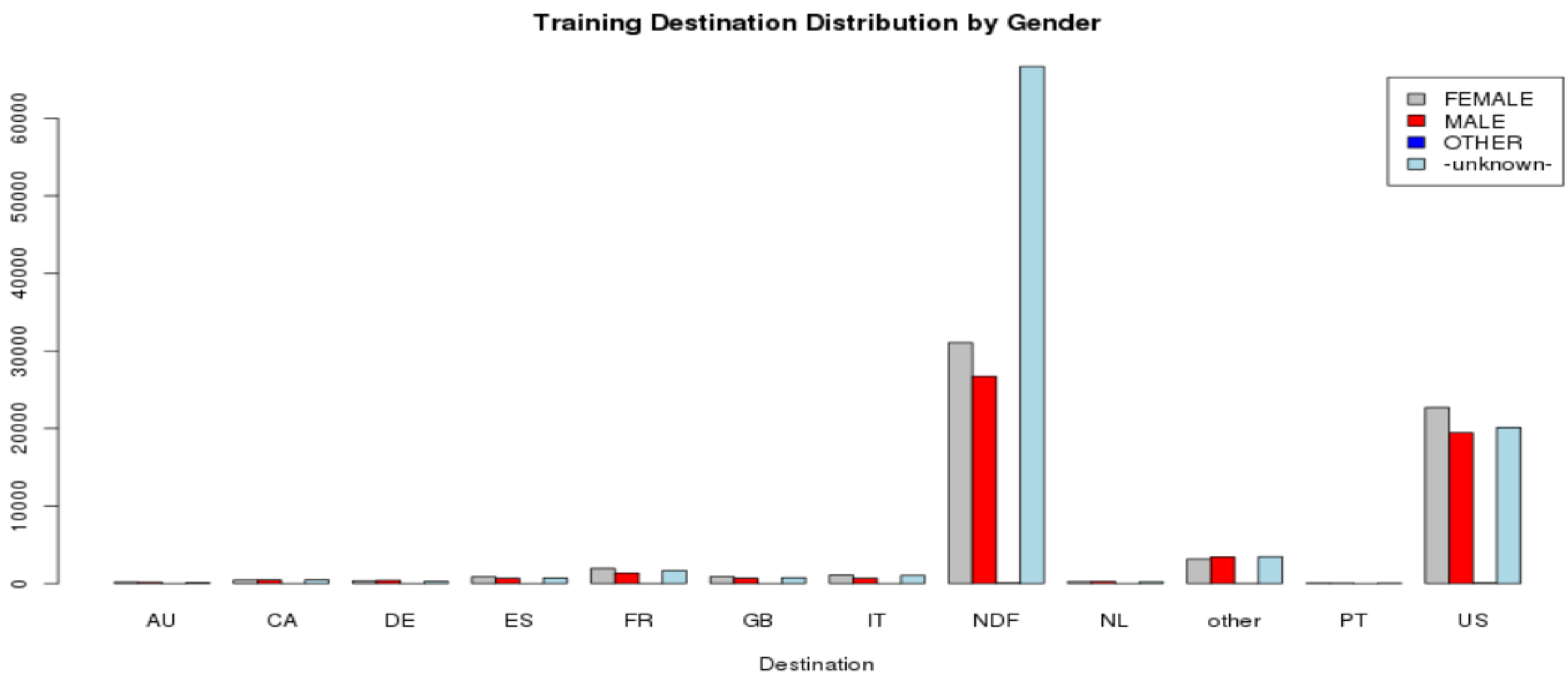


Figure 5-2. Confusion Matrix of Baseline(Decision Tree) and Final Model(Random Forest)

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	<-classified as
3							219				210	(a): class AU
	9						619		2		513	(b): class CA
	1	13					466		2		367	(c): class DE
	1	1	25	1			1021		1		750	(d): class ES
1		1	1	44		2	2214				1756	(e): class FR
			4	1	12		1022		2		819	(f): class GB
	1	2	2	4		24	1281		4		950	(g): class IT
	1	3	2	11	1	4	86150		12		13451	(h): class NDF
			1	3		2	342	1	1		260	(i): class NL
1	2	1	5	8	2	5	4558	1	93		3400	(j): class other
				1			105		2		66	(k): class PT
1	2	3	3	9	6	5	25855		24		23993	(l): class US

predict actual	AU	CA	DE	ES	FR	GB	IT	NDF	NL	PT	US	other
AU	488	0	0	0	2	0	0	41	0	0	8	0
CA	0	1211	0	2	4	1	1	176	1	1	27	4
DE	0	1	933	0	0	0	2	98	0	0	25	2
ES	0	1	0	1926	1	0	1	266	0	0	51	3
FR	1	1	0	5	4294	1	2	590	3	0	116	10
GB	0	2	1	0	0	1972	0	287	0	0	56	6
IT	0	1	0	2	2	4	2347	395	1	0	75	8
NDF	7	17	6	28	68	33	47	122574	10	3	1579	171
NL	0	0	0	0	2	0	0	88	653	0	16	3
PT	0	0	0	0	0	2	0	23	0	186	5	1
US	6	7	9	24	43	24	27	6303	8	2	55816	107
other	1	2	3	6	15	2	7	1353	3	0	274	8428

Appendix

Figure 6-1. Averaged Sanitation Indexes for Total and Four Income Subgroups of 158 Countries (1990-2015)

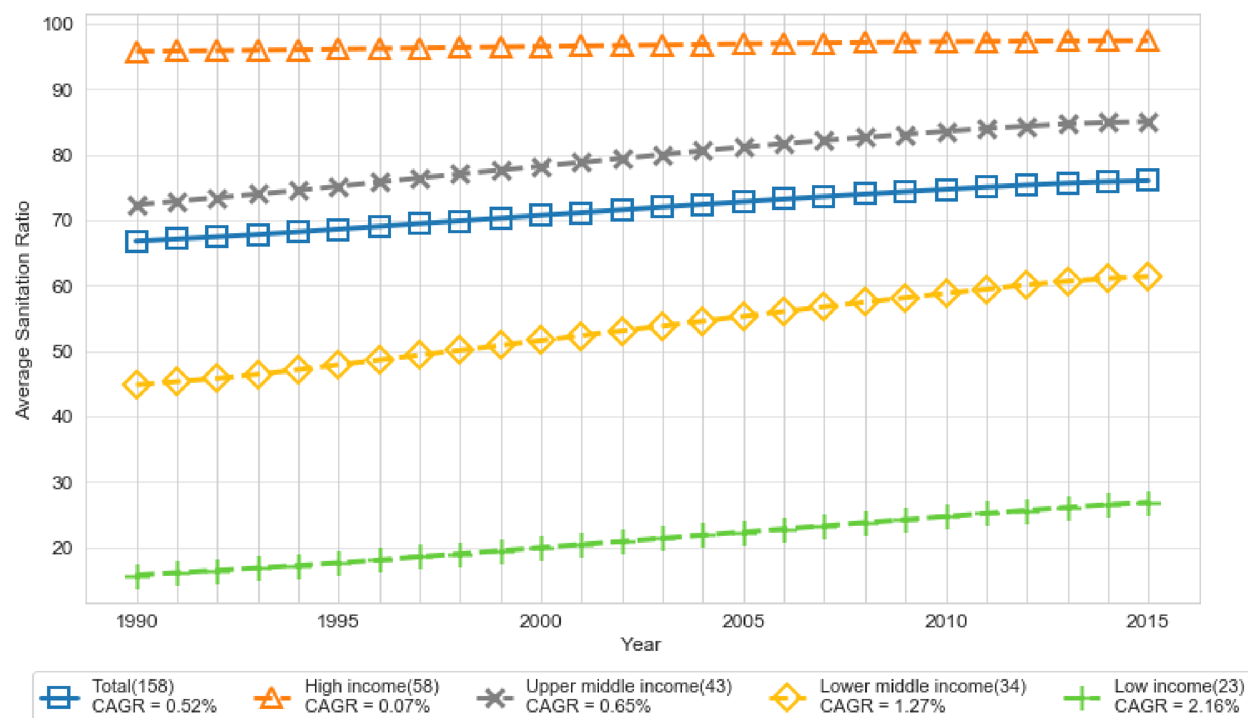
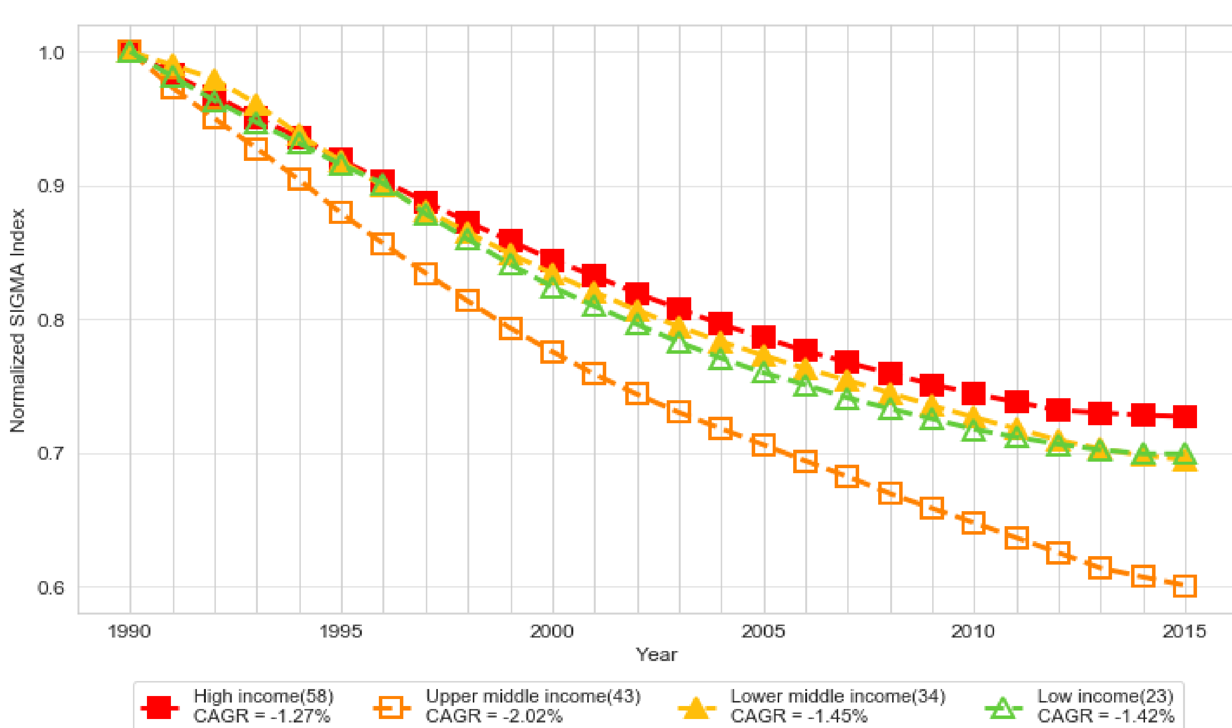


Figure 6-2. Equation of Sigma and Normalized Sigma Indexes for Four Income Subgroups (1990-2015)



$$std^n = \sqrt{\frac{1}{m-1} * \sum_{i=1}^{m} (X_i^n - \bar{X}^n)^2}$$

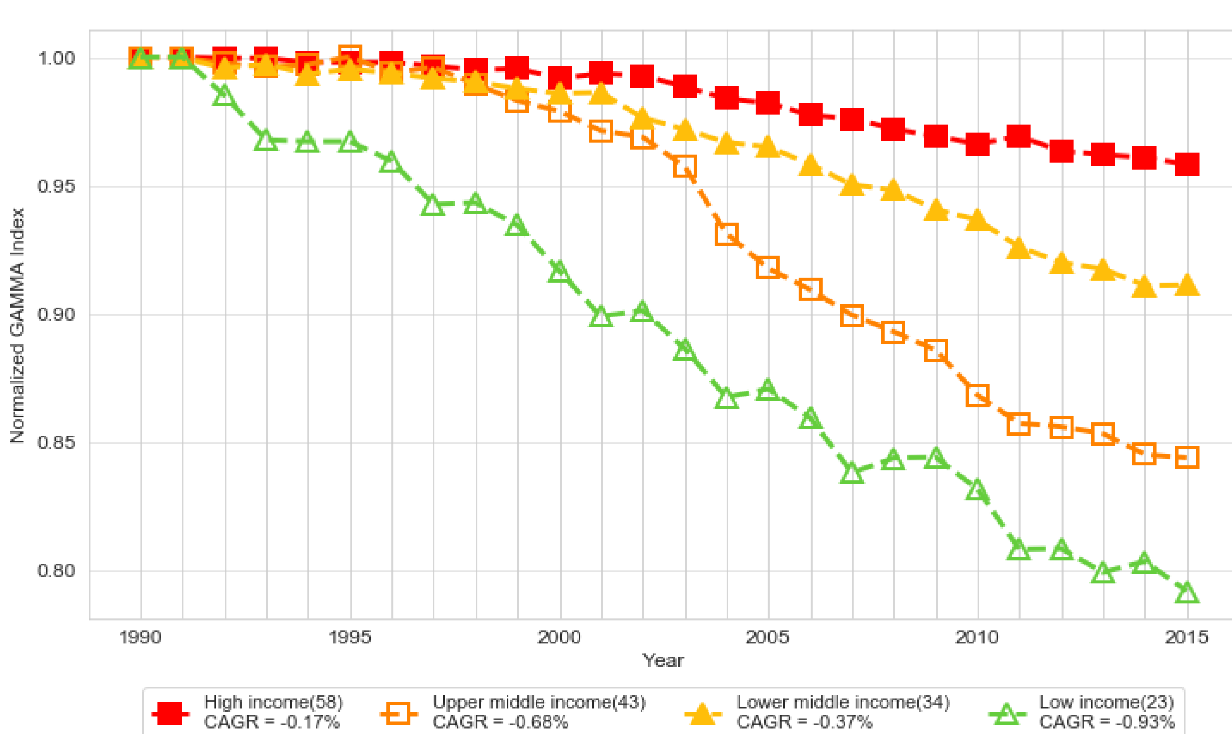
$$\bar{X}^n = \frac{1}{m} * \sum_{i=1}^{m} X_i^n$$

$$CV^n = std^n \div \bar{X}^n$$

$$\sigma^n = CV^n \div CV^{n_0}$$

$$if\ n = n_0, \qquad \sigma^n = 1$$

Figure 6-3. Equation of Gamma and Normalized Gamma Indexes for Four Income Subgroups (1990-2015)



$$\gamma^n = \frac{var(Rank_{i=1}^m(X_i^n) + Rank_{i=1}^m(X_i^{n_0}))}{4 * var(Rank_{i=1}^m(X_i^{n_0}))}$$

$$m = number\ of\ countries$$

$$n = Each\ year, \qquad k = number\ of\ years$$

$$var(X_i^n) = \frac{1}{m-1} * \sum_{i=1}^{m} (X_i^n - \overline{X^n})^2$$

Appendix

Table 6-1. Normalized Sigma and Gamma Sanitation Indexes of Total and Four Income Subgroups (1990-2015)

	All Countries (158)		High (58)		Upper Middle (43)		Lower Middle (34)		Low (23)	
Year	Sigma	Gamma	Sigma	Gamma	Sigma	Gamma	Sigma	Gamma	Sigma	Gamma
1990	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1991	0.9901	0.9997***	0.9834	1.0000***	0.9732	0.9998***	0.9900	0.9997***	0.9815	1.0000***
1992	0.9806	0.9990***	0.9672	0.9996***	0.9500	0.9975***	0.9794	0.9956***	0.9635	0.9851***
1993	0.9693	0.9988***	0.9503	0.9996***	0.9276	0.9962***	0.9610	0.9972***	0.9470	0.9677***
1994	0.9568	0.9989***	0.9359	0.9979***	0.9042	0.9971***	0.9378	0.9934***	0.9315	0.9671***
1995	0.9444	0.9991***	0.9191	0.9980***	0.8792	1.0001***	0.9183	0.9953***	0.9155	0.9671***
1996	0.9324	0.9988***	0.9034	0.9979***	0.8564	0.9939***	0.8991	0.9937***	0.9013	0.9595***
1997	0.9199	0.9983***	0.8877	0.9966***	0.8343	0.9960***	0.8814	0.9917***	0.8787	0.9426***
1998	0.9082	0.9989***	0.8728	0.9950***	0.8135	0.9895***	0.8649	0.9904***	0.8597	0.9432***
1999	0.8966	0.9977***	0.8584	0.9955***	0.7933	0.9830***	0.8489	0.9878***	0.8410	0.9345***
2000	0.8855	0.9968***	0.8443	0.9918***	0.7757	0.9789***	0.8341	0.9858***	0.8243	0.9167***
2001	0.8747	0.9971***	0.8322	0.9935***	0.7587*	0.9713***	0.8203	0.9862***	0.8097	0.8990***
2002	0.8643	0.9970***	0.8194	0.9928***	0.7438*	0.9688***	0.8067	0.9763***	0.7961	0.9012***
2003	0.8543*	0.9954***	0.8076	0.9885***	0.7302*	0.9574***	0.7944	0.9719***	0.7828	0.8864***
2004	0.8446*	0.9922***	0.7960*	0.9838***	0.7180**	0.9312***	0.7830	0.9667***	0.7707	0.8673***
2005	0.8354*	0.9910***	0.7861*	0.9822***	0.7060**	0.9176***	0.7729	0.9650***	0.7599	0.8706***
2006	0.8263**	0.9892***	0.7761*	0.9774***	0.6939**	0.9094***	0.7631	0.9583***	0.7507	0.8594***
2007	0.8175**	0.9859***	0.7677**	0.9760***	0.6822**	0.8994***	0.7542	0.9502***	0.7410	0.8379***
2008	0.8086**	0.9864***	0.7590**	0.9720***	0.6696**	0.8929***	0.7446	0.9481***	0.7329	0.8436***
2009	0.8001**	0.9848***	0.7511**	0.9691***	0.6584**	0.8858***	0.7354	0.9406***	0.7254	0.8439***
2010	0.7918**	0.9827***	0.7443**	0.9662***	0.6476***	0.8682***	0.7264	0.9365***	0.7175	0.8315***
2011	0.7838***	0.9835***	0.7377**	0.9691***	0.6364***	0.8572***	0.7181	0.9259***	0.7117	0.8080***
2012	0.7757***	0.9824***	0.7318**	0.9633***	0.6251***	0.8559***	0.7094	0.9199***	0.7064	0.8083***
2013	0.7686***	0.9812***	0.7298**	0.9620***	0.6137***	0.8532***	0.7027	0.9173***	0.7021	0.7991***
2014	0.7628***	0.9804***	0.7280**	0.9608***	0.6070***	0.8450***	0.6979*	0.9110***	0.6992	0.8034***
2015	0.7586***	0.9802***	0.7270**	0.9583***	0.6007***	0.8437***	0.6947*	0.9112***	0.6991	0.7917***
CAGR (%)	-1.10%	-0.08%	-1.27%	-0.17%	-2.02%	-0.68%	-1.45%	-0.37%	-1.42%	-0.93%

*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

Appendix

Table 7-1. Book of delta hedging portfolio P&L with daily, cumulative, Greeks

날짜	실제 손익				Greeks Profit & Loss			
	DAILY			누적 손익	Daily Greeks	Gamma 손익	Theta 손익	Vega 손익
	옵션 손익	선물 손익	손익					
2017.11.10	₩ -	₩ -	₩ -	₩ -	₩ -	₩ -	₩ -	₩ -
2017.11.13	-₩ 82,940,000	₩ 68,175,000	-₩ 14,765,000	-₩ 14,765,000	-₩ 34,608,194	-₩ 10,229,487	₩ 35,068,449	-₩ 59,447,157
2017.11.14	₩ 65,870,000	₩ 15,375,000	₩ 81,245,000	₩ 66,480,000	₩ 26,019,513	-₩ 2,227,956	₩ 13,241,956	₩ 15,005,514
2017.11.15	-₩ 224,990,000	₩ 144,587,500	-₩ 80,402,500	-₩ 13,922,500	-₩ 31,046,395	-₩ 9,826,873	₩ 13,966,703	-₩ 35,186,225
2017.11.16	₩ 243,540,000	-₩ 173,075,000	₩ 70,465,000	₩ 56,542,500	₩ 67,685,166	-₩ 24,871,429	₩ 25,399,807	₩ 67,156,789
2017.11.17	-₩ 5,130,000	₩ 20,475,000	₩ 15,345,000	₩ 71,887,500	₩ 16,564,104	-₩ 144,184	₩ 13,473,300	₩ 3,234,988
2017.11.20	-₩ 83,820,000	₩ 88,875,000	₩ 5,055,000	₩ 76,942,500	-₩ 24,045,784	-₩ 7,188,146	₩ 22,784,125	-₩ 39,641,763
2017.11.21	₩ 133,640,000	-₩ 95,550,000	₩ 38,090,000	₩ 115,032,500	₩ 33,704,130	-₩ 1,755,314	₩ 15,095,750	₩ 20,363,694
2017.11.22	₩ 105,810,000	-₩ 104,550,000	₩ 1,260,000	₩ 116,292,500	-₩ 34,860	-₩ 18,887,586	₩ 14,527,200	₩ 4,325,526
2017.11.23	-₩ 56,750,000	₩ 47,025,000	-₩ 9,725,000	₩ 106,567,500	₩ 6,713,021	-₩ 3,858,558	₩ 12,314,900	-₩ 1,743,321
2017.11.24	₩ 41,490,000	-₩ 40,000,000	₩ 1,490,000	₩ 108,057,500	₩ 19,131,648	-₩ 3,799,212	₩ 22,324,708	₩ 606,152
2017.11.27	-₩ 421,450,000	₩ 260,350,000	-₩ 161,100,000	-₩ 53,042,500	-₩ 153,700,556	-₩ 139,253,702	₩ 36,625,950	-₩ 51,072,804
2017.11.28	₩ 202,900,000	-₩ 179,550,000	₩ 23,350,000	-₩ 29,692,500	₩ 39,512,040	-₩ 14,529,471	₩ 19,754,675	₩ 34,286,836
2017.11.29	₩ 9,290,000	₩ -	₩ 9,290,000	-₩ 20,402,500	-₩ 972,604	-₩ 613,317	₩ 20,142,500	-₩ 17,261,190
2017.11.30	-₩ 684,770,000	₩ 457,800,000	-₩ 228,332,500	-₩ 247,372,500	-₩ 212,058,608	-₩ 183,640,077	₩ 32,800,667	-₩ 61,219,198
2017.12.01	₩ 6,740,000	₩ 66,500,000	₩ 73,240,000	-₩ 174,132,500	₩ 36,710,446	-₩ 1,939,969	₩ 43,488,542	-₩ 4,838,127
2017.12.04	₩ 477,530,000	-₩ 549,675,000	-₩ 72,145,000	-₩ 246,277,500	-₩ 91,702,825	-₩ 140,606,447	₩ 46,562,375	₩ 2,341,246
2017.12.05	₩ 180,040,000	-₩ 162,800,000	₩ 17,240,000	-₩ 229,037,500	₩ 35,155,877	-₩ 5,624,660	₩ 26,602,550	₩ 14,177,988
2017.12.06	-₩ 580,720,000	₩ 429,412,500	-₩ 151,307,500	-₩ 380,345,000	-₩ 193,376,626	-₩ 189,778,591	₩ 25,339,125	-₩ 28,937,160
2017.12.07	-₩ 128,470,000	₩ 172,900,000	₩ 44,430,000	-₩ 335,915,000	₩ 44,407,402	-₩ 9,294,440	₩ 58,640,000	-₩ 4,938,158
2017.12.08	₩ 232,350,000	-₩ 142,100,000	₩ 90,250,000	-₩ 245,665,000	₩ 83,326,125	-₩ 12,808,564	₩ 61,312,792	₩ 34,821,897
2017.12.11	₩ 90,400,000	-₩ 18,025,000	₩ 72,375,000	-₩ 173,290,000	₩ 73,917,259	-₩ 1,978,877	₩ 68,083,250	₩ 7,812,886
2017.12.12	-₩ 127,600,000	₩ 185,325,000	₩ 57,725,000	-₩ 115,565,000	₩ 37,645,121	-₩ 14,591,962	₩ 49,476,450	₩ 2,760,634
2017.12.13	₩ 545,020,000	-₩ 499,800,000	₩ 45,220,000	-₩ 70,345,000	-₩ 6,910,066	-₩ 58,916,714	₩ 45,670,550	₩ 6,336,098
2017.12.14	₩ 309,270,000	-₩ 227,650,000	₩ 81,470,000	₩ 11,125,000	₩ 13,545,040	-₩ 28,369,291	₩ 44,419,300	-₩ 2,504,969

Figure 7-2. Comparison for Greeks and Daily P&L by time

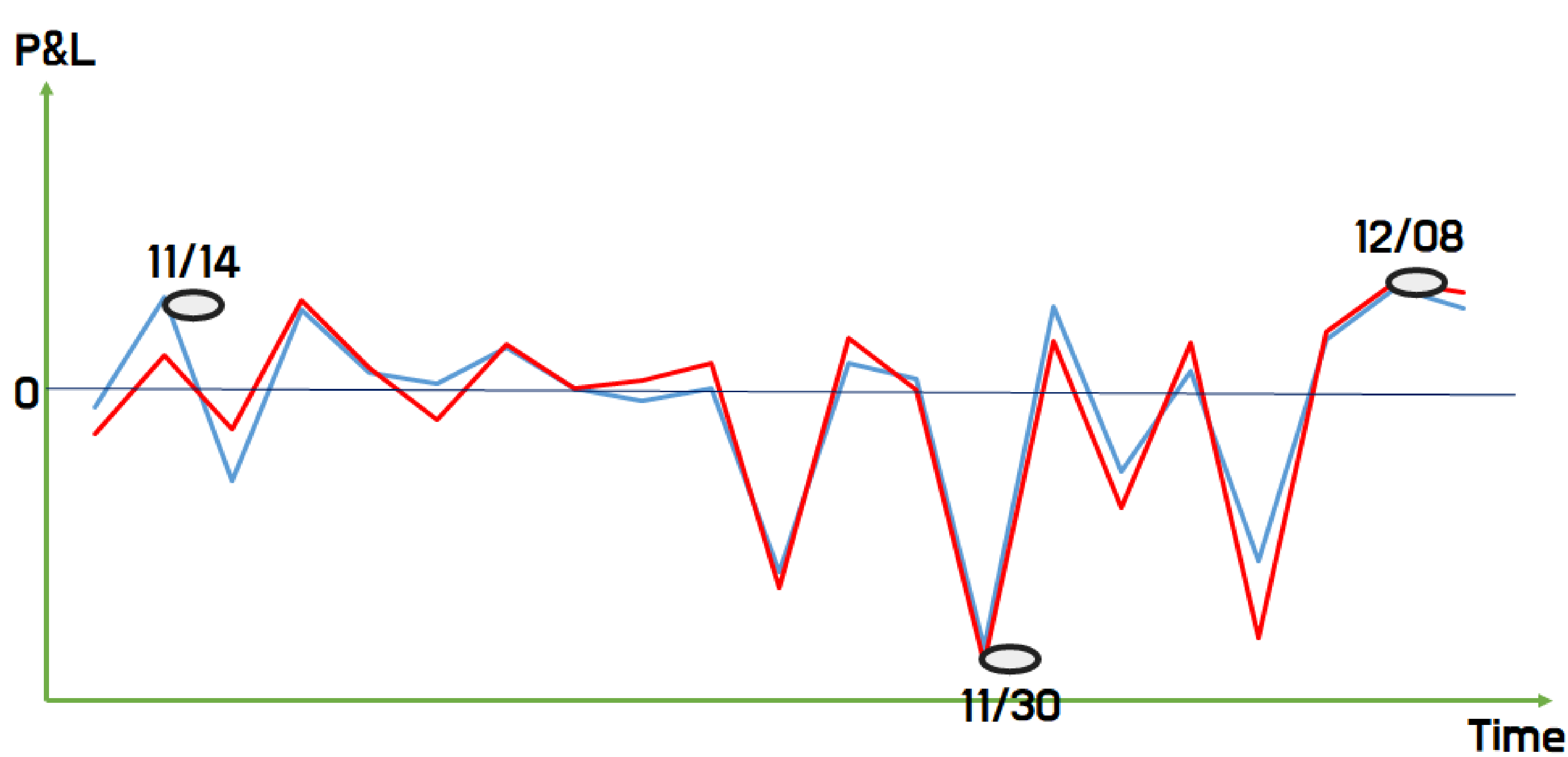


Figure : Greeks, Daily P&L 차트

Greeks P&L Daily P&L

Appendix

Figure 7-3. Equation of Greeks and Taylor Expansion

Greeks

- Delta
  - $\Delta_c = N(d_1) > 0, \qquad \Delta_p = -N(-d_1) < 0$
- Gamma
  - $\Gamma_c = \dfrac{\partial^2 C}{\partial S^2} = N'(d_1)\dfrac{\partial d_1}{\partial S} = \dfrac{N'(d_1)}{S\sigma\sqrt{T-t}} = \dfrac{\partial^2 P}{\partial S^2} = \Gamma_p$
- Theta
  - $\Theta_c = \dfrac{\partial C}{\partial t} = -\dfrac{S\sigma N'(d_1)}{2\sqrt{T-t}} - r{\cdot}Ke^{-r(T-t)}N(d_2) < 0$
  - $\Theta_p = \dfrac{\partial P}{\partial t} = -\dfrac{S\sigma N'(d_1)}{2\sqrt{T-t}} + r{\cdot}Ke^{-r(T-t)}N(-d_2)$
- Vega
  - $\nu_c = \dfrac{\partial C}{\partial \sigma} = S\sqrt{T-t}N(d_1) = \dfrac{\partial P}{\partial \sigma} = \nu_p > 0$
- Rho
  - $\rho_c = \dfrac{\partial C}{\partial r} = (T-t)K^{-r(T-t)}N(d_2) > 0$
  - $\rho_p = \dfrac{\partial P}{\partial r} = -(T-t)K^{-r(T-t)}N(-d_2) < 0$

Taylor Expansion

$$\Delta f = \Theta \Delta t + \frac{1}{2}\Gamma(\Delta S)^2 + \nu \Delta \sigma$$

Appendix

Figure 8-1. Optima Portfolio

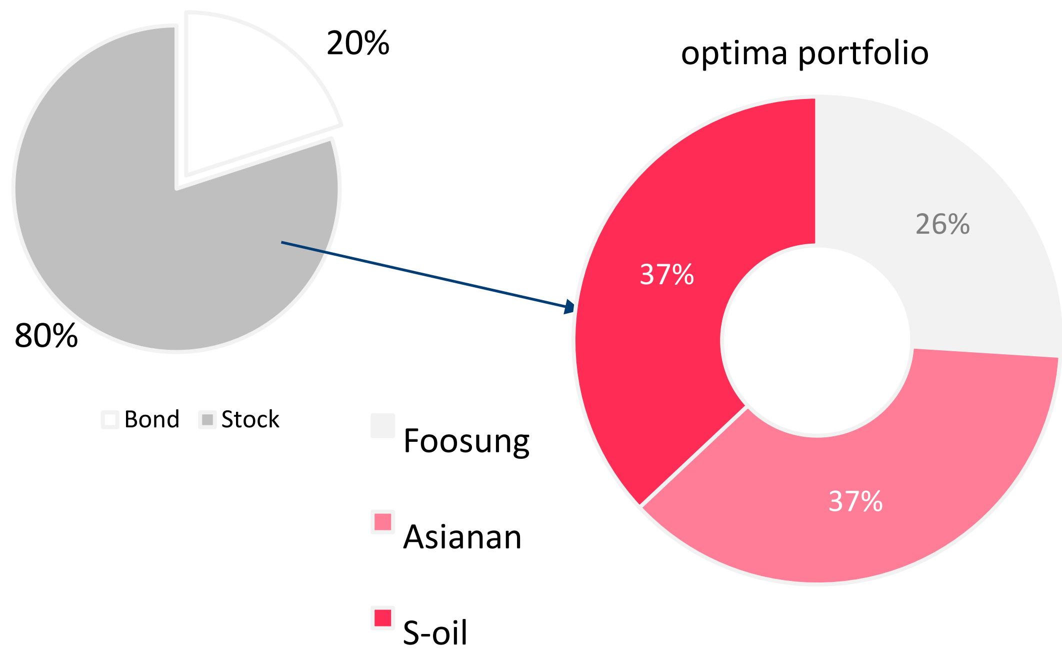
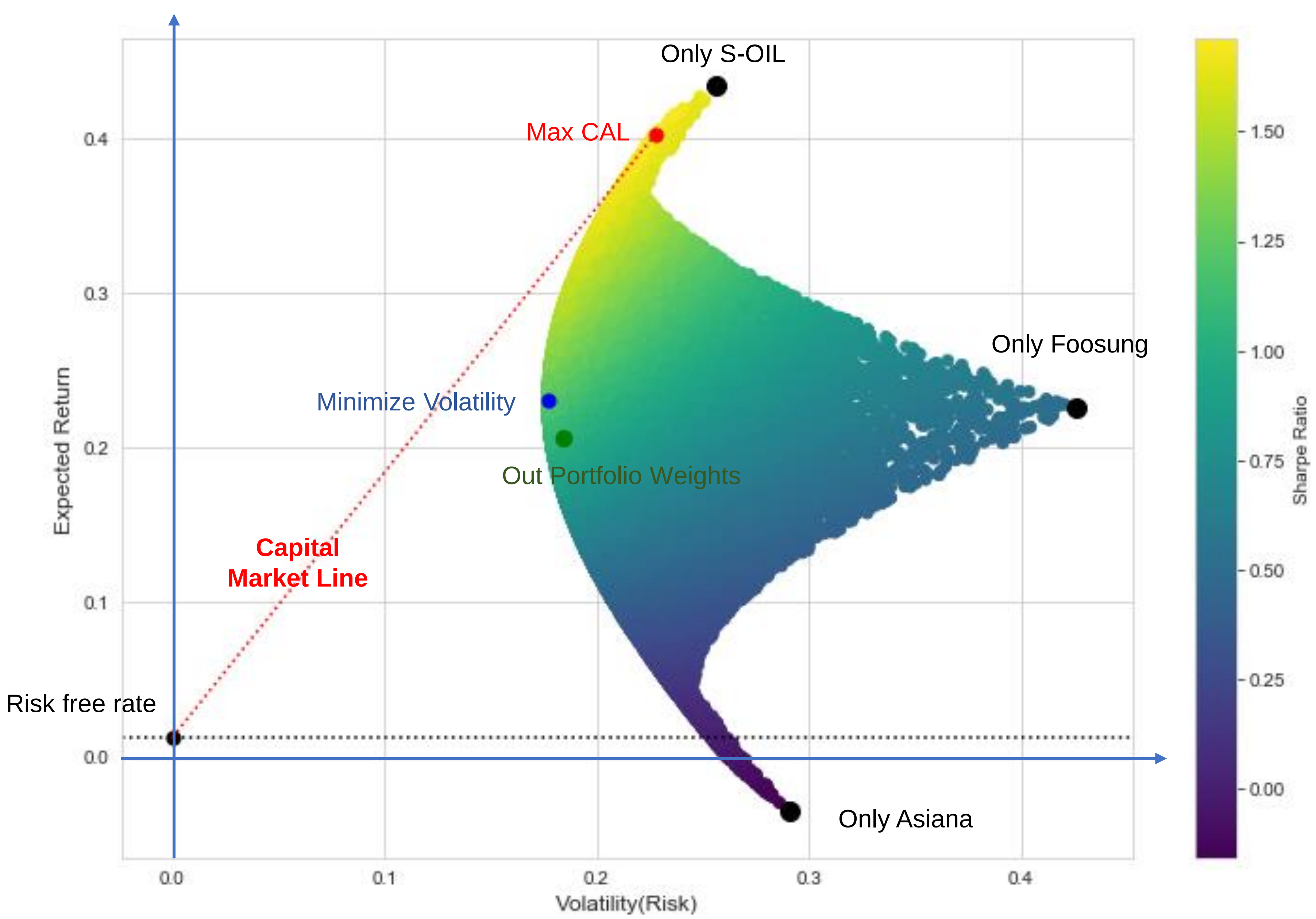


Figure 8-2. Efficient Frontier of 3 stocks and 6 weights point




```
array([[1.        , 0.        , 0.        ],   # Only Foosung
       [0.        , 1.        , 0.        ],   # Only S-OIL
       [0.        , 0.        , 1.        ],   # Only Asiana
       [0.14618618, 0.84997688, 0.00383694],   # Max CAL
       [0.1575469 , 0.47845432, 0.36399878],   # Minimize Risk
       [0.26      , 0.37      , 0.37      ]])  # Our Port.
```

Appendix

Figure 8-3. Cumulative Returns for 6 portfolio weights

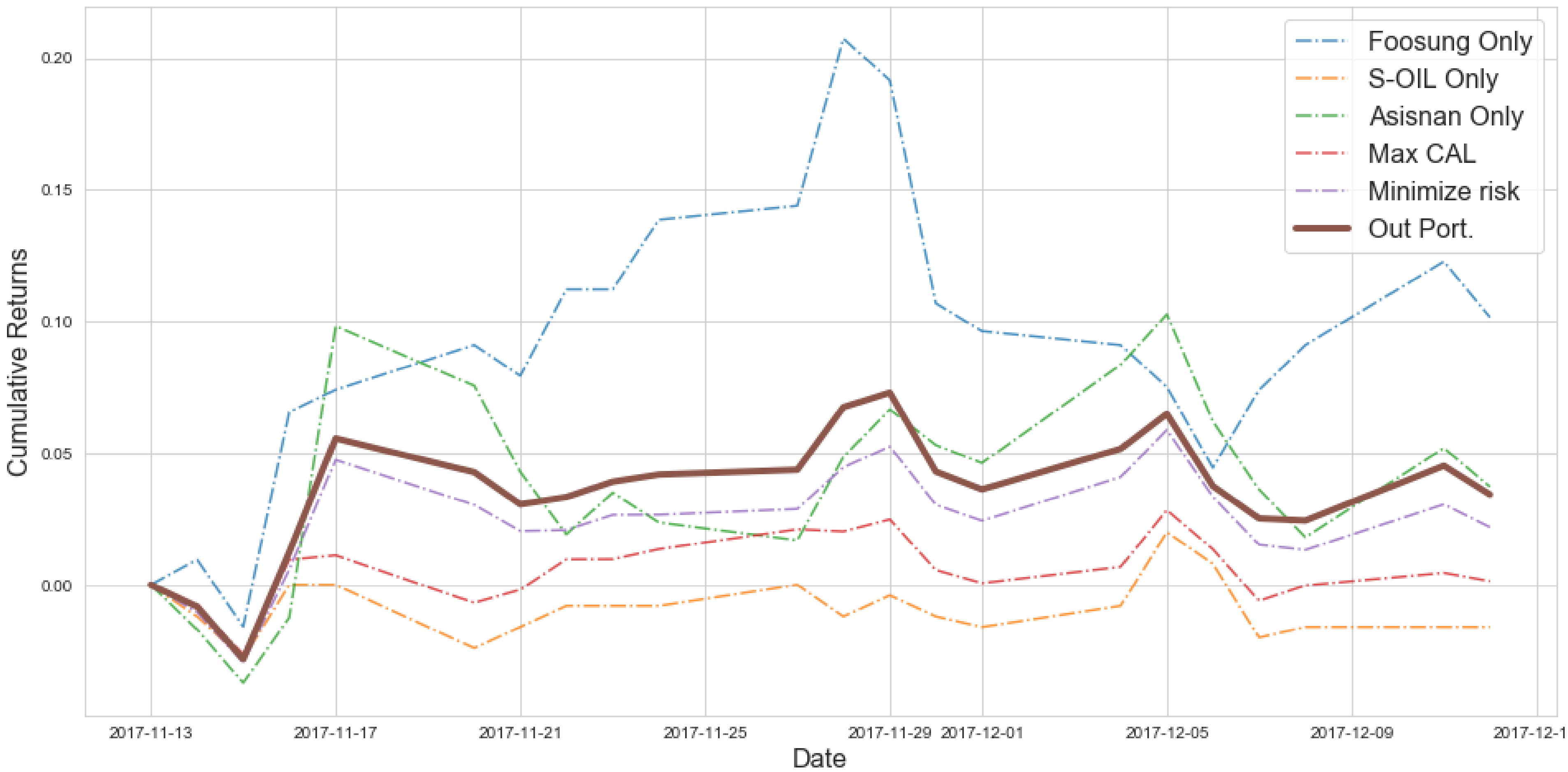


Figure 9-1. Bar Chart of portfolio weight ratio

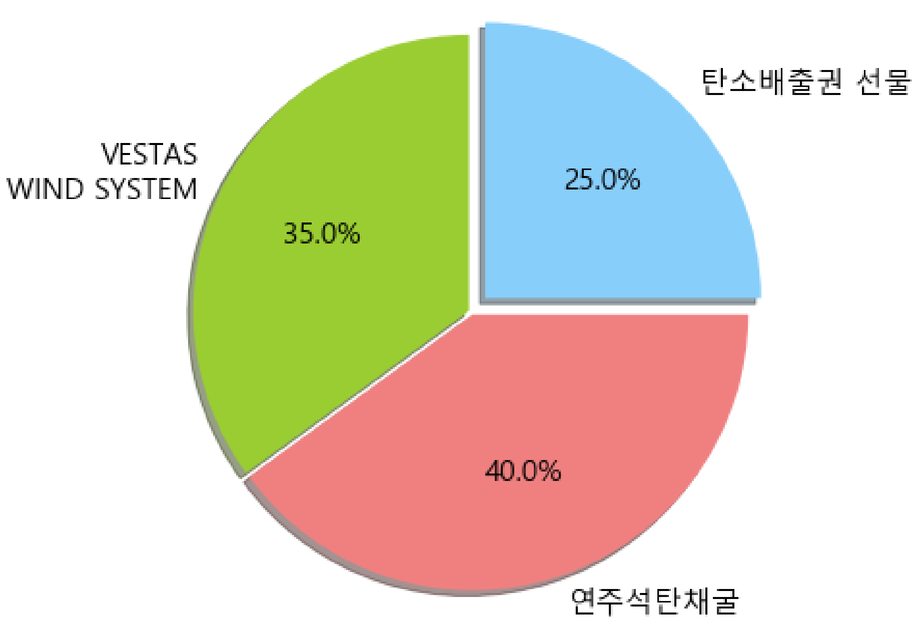


Figure 9-2. Analyze Case study of tail risk

브렉시트(2016/6/23~7/22)		
베스타스	연주석탄	탄소
-1%	16%	-14%
헤지된 수익률		
3%		
베스타스 급락시(2016/9/18~12/8)		
베스타스	연주석탄	탄소
-24%	-5%	40%
헤지된 수익률		
-1%		
연주 급락시 (2015/8/19-8/24)		
베스타스	연주석탄	탄소
-16%	-20%	3%
헤지된 수익률		
-13%		