변수들을 통한 데이터 분석 기법으로 미래 금값 예측
Forecast Future Gold Price with Data Analysis Techniques

오승엽, 정현우, 권현준, 이준석
명지대학교 산업경영공학과

1. 개요

연구 배경

금융위기를 겪으며 금리가 낮아지고 안전자산을 선택하는 경향이 높아짐

이에 대해서 안전자산인 금값을 예측

연구 목표

현실세계에서 금값에 미치는 변수들을 탐색, 발굴하여 결과적으로 금값에 가장 큰 영향을 미치는 변수와 미래의 금값을 예측하는 것이다.

2. 이론적 배경

관련 이론

-데이터 마이닝
: 일반적인 의미는 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아 내는 것

-선형 회귀 분석
: 회귀 분석은 관찰된 변수들에 대해 변수 사이의 모형을 구한 뒤 적합도를 측정해 내는 분석 방법

-주성분 분석
: 고차원의 데이터를 저차원의 데이터로 환원시키는 기법

-의사결정 나무
: 의사결정 규칙과 그 결과들을 트리 구조로 도식화한 의사 결정 지원 도구의 일종

3. 변수 선정

선정 이유

1. 금값의 변수 선정에 있어서 가장 중요한 요인은 금(Gold)은 수요와 공급에 따라 값이 결정되는 시장경제가 적용되는 것이다. 그렇기 때문에 대표적인 금융시장의 지표들을 변수로 선정

2. 금값은 '값'이라는 특성상 돈의 가치와 밀접한 관련이 있다고 판단하였다. 특히 기축통화인 달러를 선두로 하여 엔화, 유로화, 위안화 등등 국가들의 돈의 가치를 또한 변수로 추가

3. 금은 무형의 재산이 아닌 유형의 재산이기 때문에 금값에는 어느정도의 물류비가 차지하고 있는 영역도 있을 것이라고 판단

4. 그렇기 때문에 국제 유가, 한국 유가, 두바이 유가 등을 변수로 선정

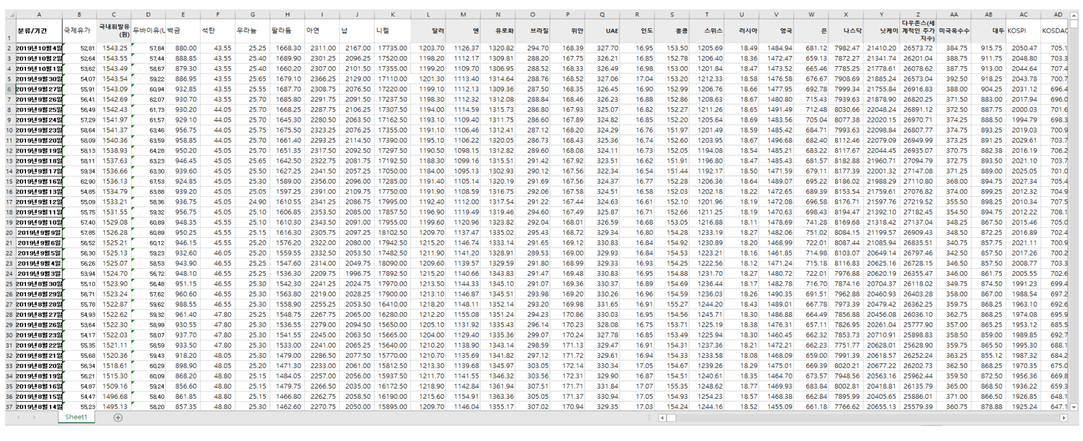
5. 마지막으로 경기 순환 이론에 의한 각종 비철금속 및 에너지 자원과 대표적 원자재의 가격을 변수로 추가

4. 설계 방법

실험 방법

실험은 데이터 분석 프로그램 중 XLMINER를 이용한다.
선정한 변수들을 입력한 뒤 위에서 언급한 기법들을 실행한다.
구체적인 실험 방법은 515개의 데이터를 Train : Validation : test의 데이터로 나누고 각각의 비율을 5 : 3 : 2 로 나누어서 분석을 진행한다.

Data Set



분류/개간	국제유가	국내휘발유(B)	두바이유(1L)_박금	석탄	우라늄	알라듐	아연	납	니켈	달러	엔	유로화	브라질	위안	UAE	인도	홍콩	스위스	러시아	영국	은	나스닥	닛케이	다우존스(세계적인 주가지수)	미국옥수수	대두	KOSPI	KOSDAQ
2019년10월4일	52.61	1543.25	57.84	880.00	43.55	25.25	1668.30	2311.00	2167.00	17735.00	1203.70	1126.37	1320.82	294.70	168.39	327.70	16.95	153.50	1205.69	18.49	1484.94	681.12	7982.47	21410.20	26573.72	384.75	915.75	2050.47	705.1
2019년10월27일	52.64	1543.55	57.44	888.85	43.55	25.40	1689.90	2301.25	2096.25	17320.00	1198.20	1112.17	1309.81	288.20	167.75	326.21	16.85	152.78	1206.40	18.36	1472.47	659.13	7872.27	21341.74	26201.04	388.75	911.75	2048.80	703.3
2019년10월18일	53.62	1543.49	58.67	879.30	43.55	25.40	1660.20	2307.00	2101.50	17355.00	1199.20	1109.70	1306.95	288.52	168.33	326.49	16.98	153.00	1201.84	18.47	1473.52	665.46	7785.25	21778.61	26078.62	387.75	913.00	2044.64	707.4
2019년9월30일	54.07	1543.54	59.22	886.95	43.55	25.65	1679.10	2366.25	2129.00	17110.00	1201.30	1113.40	1314.64	288.76	168.52	327.06	17.04	153.20	1212.33	18.58	1476.58	676.67	7908.69	21885.24	26573.04	392.50	918.25	2043.78	700.7
2019년9월27일	55.91	1543.09	60.94	932.85	43.55	25.55	1687.70	2308.75	2076.50	17220.00	1199.10	1112.13	1309.36	287.50	168.35	326.45	16.90	152.99	1206.76	18.66	1477.95	692.78	7999.34	21755.84	26916.83	388.00	904.25	2031.12	696.4
2019년9월26일	56.41	1542.69	62.07	930.70	43.55	25.70	1685.80	2291.75	2091.50	17237.50	1198.30	1112.82	1312.08	288.84	168.46	326.23	16.88	152.86	1208.63	18.67	1480.80	715.43	7939.63	21878.90	26820.25	387.50	883.00	2017.94	696.0
2019년9월25일	56.49	1542.43	61.73	930.20	44.05	25.70	1668.25	2287.75	2106.25	17307.50	1194.00	1114.59	1315.73	286.80	167.93	325.07	16.82	152.27	1211.26	18.65	1491.49	712.48	8030.66	22048.24	26891.12	372.50	887.75	2000.03	701.6
2019년9월24일	57.29	1541.97	61.57	929.10	44.05	25.70	1645.30	2280.50	2063.50	17162.50	1193.10	1109.40	1311.75	286.60	167.89	324.82	16.85	152.20	1205.64	18.69	1483.56	705.04	8077.38	22020.15	26970.71	374.25	888.50	1994.79	698.3
2019년9월23일	58.64	1541.37	63.46	956.75	44.05	25.75	1675.50	2323.25	2076.25	17355.00	1191.10	1106.46	1312.41	287.12	168.20	324.29	16.76	151.97	1201.49	18.59	1485.42	684.71	7993.63	22098.84	26807.77	374.75	893.25	2019.03	700.9
2019년9월20일	58.09	1540.36	63.59	958.85	44.05	25.70	1661.40	2293.25	2114.50	17390.00	1195.10	1106.22	1320.05	286.73	168.43	325.36	16.74	152.60	1203.95	18.67	1496.68	682.40	8112.46	22079.09	26949.99	373.25	891.25	2029.61	703.7
2019년9월19일	58.13	1538.93	64.28	950.20	45.05	25.70	1651.35	2317.50	2092.50	17297.50	1190.50	1098.15	1312.82	289.60	168.08	324.11	16.73	152.05	1194.08	18.54	1485.21	683.22	8117.67	22044.45	26935.07	370.75	882.38	2016.19	706.2
2019년9월18일	58.11	1537.63	63.23	946.45	45.05	25.65	1642.50	2322.75	2081.75	17192.50	1188.30	1099.16	1315.51	291.42	167.92	323.51	16.62	151.91	1196.80	18.47	1485.43	681.57	8182.88	21960.71	27094.79	372.75	893.50	2021.10	703.7
2019년9월17일	58.34	1536.66	63.30	939.60	45.05	25.50	1627.25	2341.50	2057.25	17050.00	1184.00	1095.13	1302.93	290.12	167.56	322.34	16.54	151.44	1192.17	18.50	1471.59	679.11	8177.39	22001.32	27147.08	371.25	889.00	2025.05	701.0
2019년9월16일	60.90	1536.13	61.53	924.85	45.05	25.30	1589.00	2356.00	2096.00	17285.00	1191.40	1105.14	1320.19	291.69	167.56	324.37	16.77	152.28	1206.36	18.64	1489.07	695.22	8186.02	21988.29	27110.80	368.00	894.75	2027.34	705.4
2019년9월13일	54.05	1534.79	63.88	939.20	45.05	25.05	1597.25	2391.00	2109.75	17750.00	1191.90	1108.59	1316.75	292.06	167.58	324.51	16.58	152.03	1202.18	18.22	1472.65	689.39	8153.54	21759.61	27076.82	374.00	899.25	2012.32	704.9
2019년9월12일	55.09	1533.21	58.36	936.75	45.05	24.90	1610.55	2341.25	2086.75	17995.00	1192.40	1112.00	1317.54	291.22	167.44	324.63	16.61	152.10	1201.96	18.19	1472.08	696.58	8176.71	21597.76	27219.52	355.50	898.25	2010.34	707.5
2019년9월11일	55.76	1531.55	59.32	956.75	45.05	25.10	1606.85	2353.50	2085.00	17857.50	1196.90	1119.49	1319.46	294.60	167.49	325.87	16.71	152.66	1211.25	18.19	1470.63	698.43	8194.47	21392.10	27182.45	354.50	894.75	2012.22	708.1
2019년9월10일	57.40	1529.08	60.89	948.35	45.55	25.10	1610.30	2343.50	2091.00	17955.00	1199.60	1120.96	1323.82	292.04	168.01	326.59	16.68	153.05	1216.88	18.11	1478.69	741.28	8169.68	21318.42	27137.04	348.25	867.50	2015.46	705.0
2019년9월9일	57.85	1526.28	60.89	950.25	45.55	25.15	1616.30	2305.75	2097.25	18102.50	1209.70	1137.47	1335.02	295.43	168.72	329.34	16.80	154.28	1233.19	18.27	1482.06	751.02	8084.15	21199.57	26909.43	348.50	872.25	2016.89	702.4
2019년9월6일	56.52	1525.21	60.12	946.15	45.55	25.20	1576.20	2322.00	2080.00	17942.50	1215.20	1146.74	1333.14	291.65	169.12	330.83	16.84	154.92	1230.89	18.20	1468.99	722.01	8087.44	21085.94	26835.51	340.75	857.75	2021.11	700.9
2019년9월5일	56.30	1525.13	59.23	932.60	46.05	25.20	1559.55	2332.50	2053.50	17482.50	1211.90	1141.20	1328.91	289.53	169.00	329.93	16.84	154.53	1223.21	18.16	1461.85	714.98	8103.07	20649.14	26797.46	342.50	857.50	2017.26	700.2
2019년9월4일	56.26	1525.07	58.53	943.90	46.55	25.25	1547.60	2314.00	2049.75	18090.00	1209.60	1139.57	1329.59	291.80	168.99	329.33	16.93	154.25	1222.56	18.12	1471.24	715.18	8116.83	20625.16	26728.15	346.50	857.50	2008.77	703.8
2019년9월3일	53.94	1524.70	56.72	948.10	46.55	25.25	1536.30	2209.75	1996.75	17892.50	1215.20	1140.66	1343.83	291.47	169.48	330.83	16.95	154.88	1231.70	18.27	1480.72	722.01	7976.88	20620.19	26355.47	346.00	861.75	2005.55	702.6
2019년8월30일	55.10	1523.90	56.48	951.15	46.55	25.30	1542.30	2241.25	2024.75	17970.00	1213.50	1144.33	1345.10	291.07	169.36	330.37	16.89	154.69	1236.44	18.17	1482.78	716.70	7874.16	20704.37	26118.02	349.75	874.50	1991.23	699.4
2019년8월29일	56.11	1523.24	57.62	960.60	46.55	25.30	1563.80	2219.00	2028.25	17900.00	1213.10	1146.87	1345.51	293.98	169.20	330.26	16.96	154.59	1236.03	18.26	1490.35	691.51	7962.88	20460.93	26403.28	358.00	867.00	1988.54	697.2
2019년8월28일	55.78	1522.87	59.62	988.55	46.55	25.30	1558.90	2255.25	2053.50	16410.00	1218.20	1148.11	1352.14	293.20	169.98	331.65	16.91	155.27	1244.20	18.43	1489.01	667.78	7973.39	20479.42	26362.25	359.75	868.25	1963.10	692.6
2019년8월27일	54.93	1522.62	59.32	961.40	47.80	25.25	1548.75	2267.75	2065.00	16280.00	1212.20	1155.08	1351.24	294.23	170.86	330.03	16.95	154.56	1245.71	18.30	1486.88	664.49	7856.88	20456.08	26036.10	362.75	868.25	1974.08	695.9
2019년8월26일	54.99	1522.01	58.99	930.55	47.80	25.30	1536.55	2279.00	2094.50	15650.00	1205.10	1131.92	1335.43	296.14	170.23	328.08	16.75	153.71	1225.19	18.38	1476.31	657.11	7826.95	20261.04	25777.90	357.00	865.25	1953.12	685.5
2019년8월23일	54.17	1522.03	58.07	937.70	47.80	25.30	1541.55	2245.00	2063.50	15665.00	1204.00	1129.40	1335.36	299.07	170.24	327.78	16.85	153.49	1225.94	18.30	1460.45	662.32	7853.73	20710.91	25898.83	358.50	859.00	1989.85	692.7
2019년8월22일	55.35	1521.11	58.59	933.50	47.80	25.30	1533.00	2241.00	2065.25	15640.00	1210.20	1138.90	1343.14	298.59	171.13	329.47	16.91	154.31	1237.36	18.21	1472.21	662.23	7751.77	20628.01	25628.90	359.75	865.50	1995.30	688.1
2019년8월21일	55.68	1520.36	59.43	918.20	48.05	25.30	1479.00	2286.50	2077.50	15770.00	1210.70	1135.69	1341.82	297.12	171.72	329.61	16.94	154.33	1233.58	18.08	1468.09	659.00	7991.39	20618.57	26252.24	363.25	855.12	1987.32	684.2
2019년8월20일	56.34	1518.61	60.09	898.90	48.05	25.20	1471.30	2233.00	2061.00	15812.50	1213.30	1139.68	1345.97	303.05	172.14	330.34	17.05	154.67	1239.26	18.29	1475.01	669.39	8020.21	20677.22	26202.73	362.50	868.25	1970.35	675.0
2019년8월19일	56.21	1515.30	60.09	868.20	48.80	25.15	1484.05	2257.00	2056.00	15937.50	1211.70	1141.55	1346.32	303.56	172.31	329.90	16.87	154.51	1240.61	18.35	1464.70	673.57	7948.56	20563.16	25962.44	359.50	872.50	1956.36	669.8
2019년8월16일	54.97	1509.16	59.24	856.60	48.80	25.15	1479.75	2266.50	2035.00	16172.50	1218.90	1142.84	1361.94	307.51	171.71	331.84	17.07	155.35	1248.62	18.77	1469.93	683.84	8002.81	20418.81	26135.79	365.00	868.50	1936.22	659.3
2019년8월15일	54.47	1496.68	58.40	861.85	48.80	25.15	1466.80	2262.75	2058.50	16190.00	1215.60	1154.91	1363.36	305.05	171.37	330.94	17.05	154.93	1254.23	18.57	1468.38	662.84	7895.99	20405.65	25886.01	371.00	866.50	1926.85	648.1
2019년8월14일	55.23	1495.13	58.20	857.35	48.80	25.30	1462.60	2270.75	2050.00	15895.00	1209.70	1146.04	1355.17	307.02	170.94	329.35	17.03	154.24	1244.16	18.52	1455.09	661.18	7766.62	20655.13	25579.39	360.75	878.88	1925.24	647.5

Sheet1

5. 결과 / 결론

실험 결과

- 다중회귀분석 결과

Residual DF	224
R?	0.9906963
Adjusted R?	0.9893257
Std. Error Estimate	436.38469
RSS	42656679

R^2 - 0.99

평균오차 – 436.38

금값을 예측하는 회귀식

Input Variables	Coefficient
Intercept	-36968.26
WTI	-28.91743
KR_Oil(won)	9.7446006
Dubai_Oil	-12.37057
Platinum	5.0108297
Coal	-39.56231
Uranium	150.92284
Palladium	1.7402597
zinc	-1.085786
lead	0.2334721
nickel	0.3140876
USD	-129.8972
JPY	23.963357
EUR	10.69828

BRL(brazil)	-1.852639
CNY(chaina)	114.21775
INR(india)	-743.1661
HKD(hk)	973.54312
CHF(swiss)	10.96192
RUB(russia)	130.12991
GBP(england	-8.001286
Silver	27.813434
Nikkei	0.120497
Nasdaq	-0.719818
dowjones	0.1880857
USA_Corn	7.5587668
bean	1.2053162
Bitcoin	-0.00582
Disnep work	29.948863
Cocacola	42.86362
coffee	-6.892323
meal (1 Bush	-4.079687
KOSPI	-5.217881
KOSDAQ	10.909796

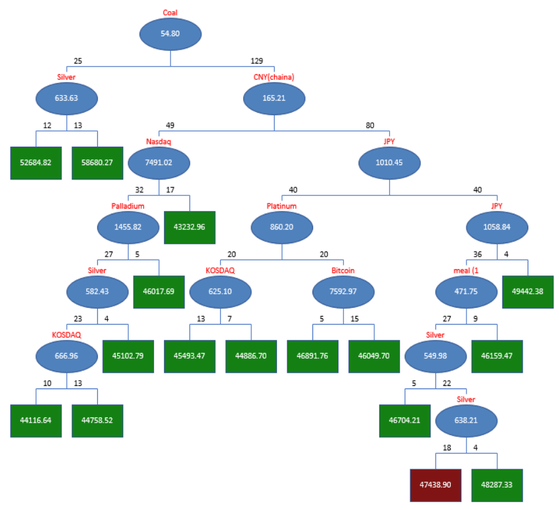
- 주성분 분석 결과

Variances	1	2	3	4	5	6	7
Variance	9479559.359	2915810.92	1110925.9	829934.5	92533.668	32827.83951	8028.4326
Variance Per	65.46371509	20.13593755	7.6718055	5.7313419	0.6390168	0.226701712	0.0554426
Cumulative V	65.46371509	85.59965264	93.271458	99.0028	99.641817	99.86851857	99.923961

Residual DF	223
R?	0.9907041
Adjusted R?	0.9893284
Std. Error Estimate	436.87012
RSS	42560778

다음과 같은 결과를 보였는데 이는 다중회귀분석과 매우 근소한 차이를 보이며 이는 Data를 랜덤으로 5:3:2로 선정하여 모델화 하였을 때 보이는 오차범위로 보여진다.
<u>결론적으로 다중회귀분석과 주성분 회귀분석의 차이는 없는 것으로 보인다.</u>

- 의사결정 나무 결과



Test Data scoring - Summary Report (Using Best Pruned Tree)

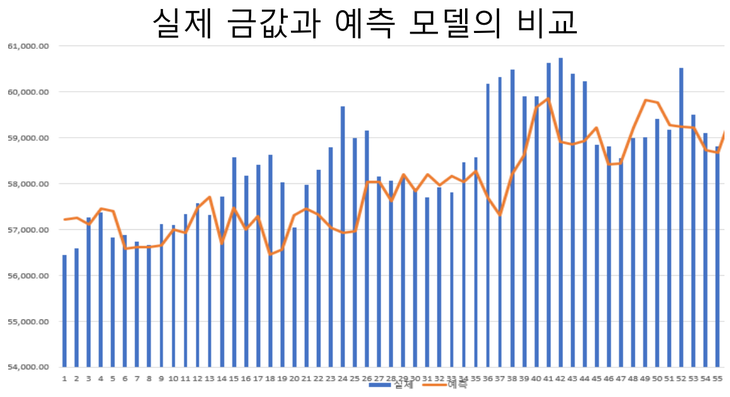
Total sum of squared errors	RMS Error	Average Error
41557183	635.19115	-24.97585229

다음과 같은 결과를 보였는데 금값이 예측치가 약 635원 정도의 평균적 오차를 보임

<u>다중회귀분석과 비교하였을 때 별도로 생각할 것도 없이 다중회귀분석이 더 우수한 것으로 보인다.</u>

결론

실제 금값과 예측 모델의 비교



*정확도 측정기준

정확한 금값이 아닌 금값이 전날에 비해 상승인지 하락인지 맞출 것(O)
예측 금값이 상승이고, 실제 금값이 상승한 경우 상승량은 관계가 없음 (O)
예측 금값이 하락이고, 실제 금값이 하락인 경우 하락량은 관계가 없음 (O)
예측 금값이 상승인데, 실제 금값이 하락한 경우 카운트 (X)
예측 금값이 하락인데, 실제 금값이 상승한 경우 카운트 (X)

- 정확도 측정결과 – 55개의 값 중 8번의 카운트
- **<u>약 85% 의 정확도</u>**

한계성

예측 모델의 명확한 한계점은 바로 연속형의 종속변수를 예측 할 때, 정확한 값을 가져도 실제 값과는 큰 차이가 있다. 현실의 많은 고려사항들을 변수화 시키기 복잡하기 때문에 이를 예측모델에 담지 못함.