Progress Seminar

: Naïve Bayes and SVM for Sentiment Classification

2018.08.02

김도연

중간보고 이후 작업

1. SNS 데이터 특성(함축어, 신조어, 은어 등)을 반영한 정규화 과정을 수행
   - 더웡, 더워, 덥니, 덥지, 덥냐, 더움, 더워라 등 -> 덥다)
2. 채널과 현상을 고려한 분류기 구축
   - 4개 채널과 4개 현상 데이터의 성격이 다르므로 각 부분집합 내에서 분류기 구축

< 채널 및 현상 구분 >

현상 / 매체	온도	강수	토지	해양
페이스북				
트위터				
인스타그램				
뉴스댓글				

NB_Model : 3개 감성 분류

Accuracy: 47.69%

	nbPred			
	1.POS	2.NEG	3.NEU	Row Total
1.POS	2876	586	1265	4727
	0.608	0.124	0.268	0.340
	0.469	0.209	0.255	
	0.207	0.042	0.091	
2.NEG	1993	1810	1755	5558
	0.359	0.326	0.316	0.400
	0.325	0.645	0.354	
	0.143	0.130	0.126	
3.NEU	1260	411	1941	3612
	0.349	0.114	0.537	0.260
	0.206	0.146	0.391	
	0.091	0.030	0.140	
Column Total	6129	2807	4961	13897
	0.441	0.202	0.357	

SVM_Model1 : 3개 감성 분류

1. 선형 커널(Linear Kernel)

Accuracy: 52.57%

```
> summary(svmModel_1)

Call:
svm(formula = target ~ ., data = trainingSet,
    kernel = "linear")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  linear
       cost:  1
      gamma:  0.006578947

Number of Support Vectors:  24081

 ( 5352 9056 9673 )


Number of Classes:  3

Levels:
 1.POS 2.NEG 3.NEU
```

	svmPred1			
	1.POS	2.NEG	3.NEU	Row Total
1.POS	2293	4656	251	7200
	0.318	0.647	0.035	0.369
	0.643	0.311	0.262	
	0.117	0.239	0.013	
2.NEG	814	7541	280	8635
	0.094	0.873	0.032	0.442
	0.228	0.503	0.292	
	0.042	0.386	0.014	
3.NEU	461	2796	428	3685
	0.125	0.759	0.116	0.189
	0.129	0.186	0.446	
	0.024	0.143	0.022	
Column Total	3568	14993	959	19520
	0.183	0.768	0.049	

SVM_Model2 : 3개 감성 분류

2. 다항식 커널(Polynomial Kernel)

Accuracy: 51.38%

```
> summary(svmModel_2)

Call:
svm(formula = target ~ ., data = trainingSet, kernel = "polynomial")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  polynomial
       cost:  1
     degree:  3
      gamma:  0.01098901
     coef.0:  0

Number of Support Vectors:  25962

 ( 10905 9716 5341 )


Number of Classes:  3

Levels:
 1.POS 2.NEG 3.NEU
```

	svmPred2			
	1.POS	2.NEG	3.NEU	Row Total
1.POS	2666	4192	342	7200
	0.370	0.582	0.048	0.369
	0.570	0.310	0.256	
	0.137	0.215	0.018	
2.NEG	1277	6863	495	8635
	0.148	0.795	0.057	0.442
	0.273	0.508	0.370	
	0.065	0.352	0.025	
3.NEU	733	2451	501	3685
	0.199	0.665	0.136	0.189
	0.157	0.181	0.374	
	0.038	0.126	0.026	
Column Total	4676	13506	1338	19520
	0.240	0.692	0.069	

SVM_Model3 : 3개 감성 분류

3. RBF커널(Radial Basis Function Kernel)

Accuracy: 54.94%

```
> summary(svmModel_3)

Call:
svm(formula = target ~ ., data = trainingSet, kernel = "radial")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.01098901

Number of Support Vectors:  25430

 ( 10713 9335 5382 )


Number of Classes:  3

Levels:
 1.POS 2.NEG 3.NEU
```

	svmPred3			
	1.POS	2.NEG	3.NEU	Row Total
1.POS	4620	2319	261	7200
	0.642	0.322	0.036	0.369
	0.507	0.251	0.222	
	0.237	0.119	0.013	
2.NEG	2784	5521	330	8635
	0.322	0.639	0.038	0.442
	0.306	0.598	0.281	
	0.143	0.283	0.017	
3.NEU	1705	1396	584	3685
	0.463	0.379	0.158	0.189
	0.187	0.151	0.497	
	0.087	0.072	0.030	
Column Total	9109	9236	1175	19520
	0.467	0.473	0.060	

SVM_Model4 : 3개 감성 분류

4. 시그모이드 커널(Sigmoid Kernel)

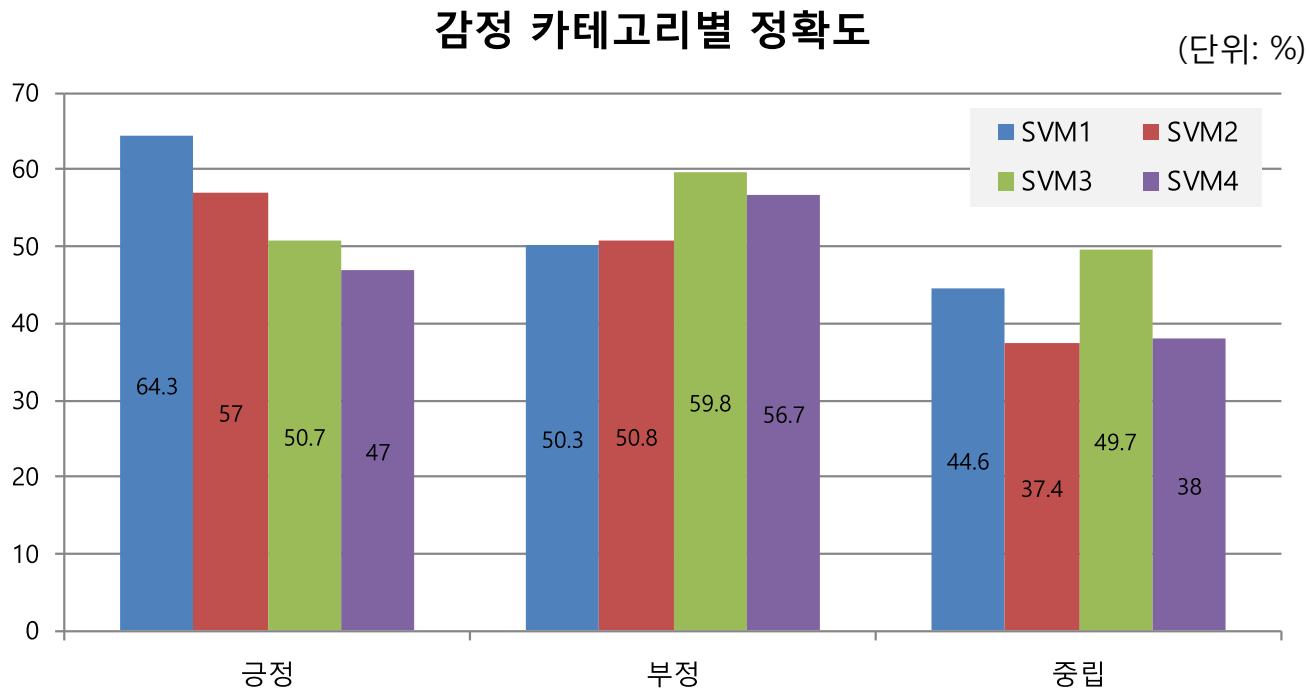
Accuracy: 50.38%

```
> summary(svmModel_4)

Call:
svm(formula = target ~ ., data = trainingSet, kernel = "sigmoid")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  sigmoid
       cost:  1
      gamma:  0.01098901
     coef.0:  0

Number of Support Vectors:  21960

 ( 8361 8310 5289 )


Number of Classes:  3

Levels:
 1.POS 2.NEG 3.NEU
```

	svmPred4			
	1.POS	2.NEG	3.NEU	Row Total
1.POS	4474	2310	416	7200
	0.621	0.321	0.058	0.369
	0.470	0.277	0.249	
	0.229	0.118	0.021	
2.NEG	3288	4725	622	8635
	0.381	0.547	0.072	0.442
	0.346	0.567	0.372	
	0.168	0.242	0.032	
3.NEU	1749	1301	635	3685
	0.475	0.353	0.172	0.189
	0.184	0.156	0.380	
	0.090	0.067	0.033	
Column Total	9511	8336	1673	19520
	0.487	0.427	0.086	

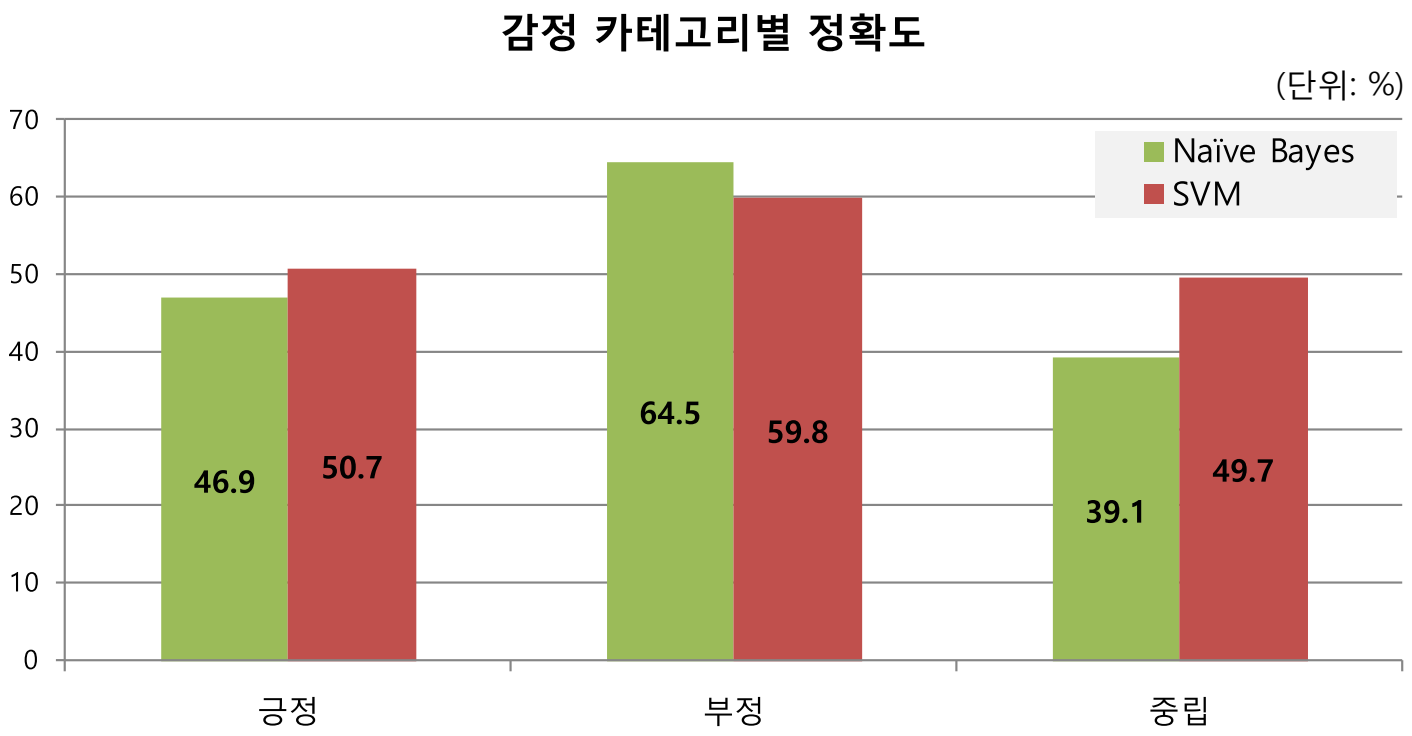
전체 데이터 : 3개 감성 class 분류 성과

Model Accuracy
- SVM1 (Linear Kernel) : 52.57%
- SVM2 (Polynomial Kernel) : 51.38%
- SVM3 (RBF Kernel) : 54.94%
- SVM4 (Sigmoid Kernel) : 50.38%



전체 데이터 : 3개 감성 class 분류 성과

Model Accuracy
- Naïve Bayes : 47.69%
- SVM(RBF Kernel) : 54.94%



NB_Model : 7개 감성 분류

Accuracy: 26.59%

	nbPred							
	1.ECS	2.EXP	3.POS	4.RAG	5.TER	6.NEG	7.NEU	Row Total
1.ECS	1144	530	83	216	133	96	315	2517
	0.455	0.211	0.033	0.086	0.053	0.038	0.125	0.129
	0.196	0.116	0.170	0.094	0.079	0.086	0.090	
	0.059	0.027	0.004	0.011	0.007	0.005	0.016	
2.EXP	571	904	41	145	105	69	313	2148
	0.266	0.421	0.019	0.068	0.049	0.032	0.146	0.110
	0.098	0.197	0.084	0.063	0.062	0.062	0.089	
	0.029	0.046	0.002	0.007	0.005	0.004	0.016	
3.POS	946	575	153	201	166	93	489	2623
	0.361	0.219	0.058	0.077	0.063	0.035	0.186	0.134
	0.162	0.125	0.314	0.088	0.098	0.083	0.139	
	0.048	0.029	0.008	0.010	0.009	0.005	0.025	
4.RAG	765	573	54	756	194	166	394	2902
	0.264	0.197	0.019	0.261	0.067	0.057	0.136	0.149
	0.131	0.125	0.111	0.330	0.115	0.148	0.112	
	0.039	0.029	0.003	0.039	0.010	0.009	0.020	
5.TER	626	574	34	293	641	130	366	2664
	0.235	0.215	0.013	0.110	0.241	0.049	0.137	0.136
	0.107	0.125	0.070	0.128	0.379	0.116	0.104	
	0.032	0.029	0.002	0.015	0.033	0.007	0.019	
6.NEG	742	694	57	389	255	428	468	3033
	0.245	0.229	0.019	0.128	0.084	0.141	0.154	0.155
	0.127	0.151	0.117	0.170	0.151	0.383	0.133	
	0.038	0.036	0.003	0.020	0.013	0.022	0.024	
7.NEU	1042	736	65	292	197	136	1165	3633
	0.287	0.203	0.018	0.080	0.054	0.037	0.321	0.186
	0.179	0.160	0.133	0.127	0.116	0.122	0.332	
	0.053	0.038	0.003	0.015	0.010	0.007	0.060	
Column Total	5836	4586	487	2292	1691	1118	3510	19520
	0.299	0.235	0.025	0.117	0.087	0.057	0.180	

SVM_Model1 : 7개 감성 분류

1. 선형 커널(Linear Kernel)

Accuracy: 30.60%

```
> summary(svmModel_1)

Call:
svm(formula = target ~ ., data = trainingSet, kernel = "linear")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  linear
       cost:  1
      gamma:  0.01098901

Number of Support Vectors:  26892

 ( 4777 4358 3579 3603 3601 3981 2993 )


Number of Classes:  7

Levels:
 1.ECS 2.EXP 3.POS 4.RAG 5.TER 6.NEG 7.NEU
```

	svmPred1							
	1.ECS	2.EXP	3.POS	4.RAG	5.TER	6.NEG	7.NEU	Row Total
1.ECS	443	56	64	138	58	211	1547	2517
	0.176	0.022	0.025	0.055	0.023	0.084	0.615	0.129
	0.389	0.120	0.126	0.078	0.051	0.084	0.129	
	0.023	0.003	0.003	0.007	0.003	0.011	0.079	
2.EXP	232	246	25	86	95	244	1220	2148
	0.108	0.115	0.012	0.040	0.044	0.114	0.568	0.110
	0.204	0.528	0.049	0.049	0.083	0.097	0.102	
	0.012	0.013	0.001	0.004	0.005	0.013	0.062	
3.POS	210	49	344	152	86	259	1523	2623
	0.080	0.019	0.131	0.058	0.033	0.099	0.581	0.134
	0.184	0.105	0.677	0.086	0.076	0.103	0.127	
	0.011	0.003	0.018	0.008	0.004	0.013	0.078	
4.RAG	78	30	21	736	123	423	1491	2902
	0.027	0.010	0.007	0.254	0.042	0.146	0.514	0.149
	0.068	0.064	0.041	0.415	0.108	0.168	0.125	
	0.004	0.002	0.001	0.038	0.006	0.022	0.076	
5.TER	40	32	21	170	557	346	1498	2664
	0.015	0.012	0.008	0.064	0.209	0.130	0.562	0.136
	0.035	0.069	0.041	0.096	0.489	0.137	0.125	
	0.002	0.002	0.001	0.009	0.029	0.018	0.077	
6.NEG	76	36	20	285	128	721	1767	3033
	0.025	0.012	0.007	0.094	0.042	0.238	0.583	0.155
	0.067	0.077	0.039	0.161	0.112	0.286	0.148	
	0.004	0.002	0.001	0.015	0.007	0.037	0.091	
7.NEU	61	17	13	206	92	318	2926	3633
	0.017	0.005	0.004	0.057	0.025	0.088	0.805	0.186
	0.054	0.036	0.026	0.116	0.081	0.126	0.244	
	0.003	0.001	0.001	0.011	0.005	0.016	0.150	
Column Total	1140	466	508	1773	1139	2522	11972	19520
	0.058	0.024	0.026	0.091	0.058	0.129	0.613	

SVM_Model2 : 7개 감성 분류

2. 다항식 커널(Polynomial Kernel)

Accuracy: 27.81%

```
> summary(svmModel_2)

Call:
svm(formula = target ~ ., data = trainingSet, kernel = "polynomial")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  polynomial
       cost:  1
     degree:  3
      gamma:  0.01098901
     coef.0:  0

Number of Support Vectors:  27638

 ( 5019 4414 3651 3698 3647 4131 3078 )


Number of Classes:  7

Levels:
 1.ECS 2.EXP 3.POS 4.RAG 5.TER 6.NEG 7.NEU
```

	svmPred2							
	1.ECS	2.EXP	3.POS	4.RAG	5.TER	6.NEG	7.NEU	Row Total
1.ECS	307	59	83	78	78	114	1798	2517
	0.122	0.023	0.033	0.031	0.031	0.045	0.714	0.129
	0.344	0.083	0.103	0.072	0.061	0.085	0.134	
	0.016	0.003	0.004	0.004	0.004	0.006	0.092	
2.EXP	156	266	61	36	106	116	1407	2148
	0.073	0.124	0.028	0.017	0.049	0.054	0.655	0.110
	0.175	0.376	0.076	0.033	0.083	0.087	0.105	
	0.008	0.014	0.003	0.002	0.005	0.006	0.072	
3.POS	160	79	389	83	118	119	1675	2623
	0.061	0.030	0.148	0.032	0.045	0.045	0.639	0.134
	0.179	0.112	0.484	0.077	0.092	0.089	0.125	
	0.008	0.004	0.020	0.004	0.006	0.006	0.086	
4.RAG	80	77	72	484	151	215	1823	2902
	0.028	0.027	0.025	0.167	0.052	0.074	0.628	0.149
	0.090	0.109	0.090	0.450	0.118	0.161	0.136	
	0.004	0.004	0.004	0.025	0.008	0.011	0.093	
5.TER	45	79	52	104	549	185	1650	2664
	0.017	0.030	0.020	0.039	0.206	0.069	0.619	0.136
	0.050	0.112	0.065	0.097	0.430	0.138	0.123	
	0.002	0.004	0.003	0.005	0.028	0.009	0.085	
6.NEG	70	78	78	172	149	422	2064	3033
	0.023	0.026	0.026	0.057	0.049	0.139	0.681	0.155
	0.078	0.110	0.097	0.160	0.117	0.316	0.154	
	0.004	0.004	0.004	0.009	0.008	0.022	0.106	
7.NEU	75	70	68	119	125	165	3011	3633
	0.021	0.019	0.019	0.033	0.034	0.045	0.829	0.186
	0.084	0.099	0.085	0.111	0.098	0.124	0.224	
	0.004	0.004	0.003	0.006	0.006	0.008	0.154	
Column Total	893	708	803	1076	1276	1336	13428	19520
	0.046	0.036	0.041	0.055	0.065	0.068	0.688	

SVM_Model3 : 7개 감성 분류

3. RBF커널(Radial Basis Function Kernel)

Accuracy: 30.66%

```
> summary(svmModel_3)

Call:
svm(formula = target ~ ., data = trainingSet, kernel = "radial")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.01098901

Number of Support Vectors:  26993

 ( 4706 4351 3621 3590 3569 4088 3068 )


Number of Classes:  7

Levels:
 1.ECS 2.EXP 3.POS 4.RAG 5.TER 6.NEG 7.NEU
```

	svmPred3							
	1.ECS	2.EXP	3.POS	4.RAG	5.TER	6.NEG	7.NEU	Row Total
1.ECS	445	122	88	234	125	266	1237	2517
	0.177	0.048	0.035	0.093	0.050	0.106	0.491	0.129
	0.383	0.121	0.109	0.087	0.077	0.100	0.129	
	0.023	0.006	0.005	0.012	0.006	0.014	0.063	
2.EXP	200	379	74	182	144	253	916	2148
	0.093	0.176	0.034	0.085	0.067	0.118	0.426	0.110
	0.172	0.375	0.092	0.068	0.089	0.095	0.096	
	0.010	0.019	0.004	0.009	0.007	0.013	0.047	
3.POS	195	125	393	260	134	295	1221	2623
	0.074	0.048	0.150	0.099	0.051	0.112	0.465	0.134
	0.168	0.124	0.489	0.097	0.083	0.110	0.128	
	0.010	0.006	0.020	0.013	0.007	0.015	0.063	
4.RAG	88	87	63	911	178	428	1147	2902
	0.030	0.030	0.022	0.314	0.061	0.147	0.395	0.149
	0.076	0.086	0.078	0.340	0.110	0.160	0.120	
	0.005	0.004	0.003	0.047	0.009	0.022	0.059	
5.TER	63	83	49	269	651	369	1180	2664
	0.024	0.031	0.018	0.101	0.244	0.139	0.443	0.136
	0.054	0.082	0.061	0.100	0.401	0.138	0.123	
	0.003	0.004	0.003	0.014	0.033	0.019	0.060	
6.NEG	100	120	66	464	218	700	1365	3033
	0.033	0.040	0.022	0.153	0.072	0.231	0.450	0.155
	0.086	0.119	0.082	0.173	0.134	0.262	0.143	
	0.005	0.006	0.003	0.024	0.011	0.036	0.070	
7.NEU	72	95	71	357	172	361	2505	3633
	0.020	0.026	0.020	0.098	0.047	0.099	0.690	0.186
	0.062	0.094	0.088	0.133	0.106	0.135	0.262	
	0.004	0.005	0.004	0.018	0.009	0.018	0.128	
Column Total	1163	1011	804	2677	1622	2672	9571	19520
	0.060	0.052	0.041	0.137	0.083	0.137	0.490	

SVM_Model4 : 7개 감성 분류

4. 시그모이드 커널(Sigmoid Kernel)

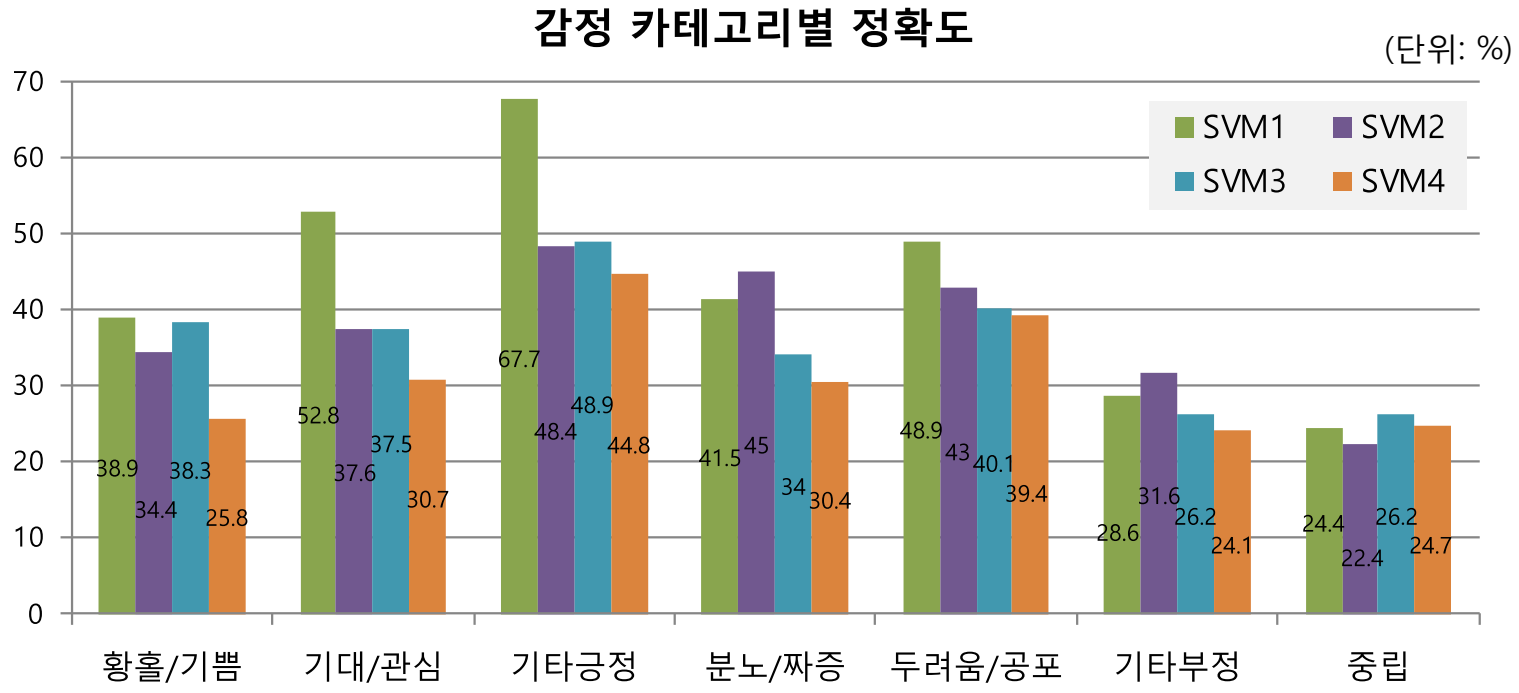
Accuracy: 27.47%

```
> summary(svmModel_4)

Call:
svm(formula = target ~ ., data = trainingSet, kernel = "sigmoid")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  sigmoid
       cost:  1
      gamma:  0.01098901
     coef.0:  0

Number of Support Vectors:  26455

 ( 4602 4349 3515 3649 3554 3786 3000 )


Number of Classes:  7

Levels:
 1.ECS 2.EXP 3.POS 4.RAG 5.TER 6.NEG 7.NEU
```

	svmPred4							
	1.ECS	2.EXP	3.POS	4.RAG	5.TER	6.NEG	7.NEU	Row Total
1.ECS	510	92	89	239	67	233	1287	2517
	0.203	0.037	0.035	0.095	0.027	0.093	0.511	0.129
	0.258	0.102	0.107	0.095	0.058	0.095	0.133	
	0.026	0.005	0.005	0.012	0.003	0.012	0.066	
2.EXP	301	277	82	188	92	249	959	2148
	0.140	0.129	0.038	0.088	0.043	0.116	0.446	0.110
	0.152	0.307	0.099	0.074	0.080	0.101	0.099	
	0.015	0.014	0.004	0.010	0.005	0.013	0.049	
3.POS	297	88	371	271	100	279	1217	2623
	0.113	0.034	0.141	0.103	0.038	0.106	0.464	0.134
	0.150	0.098	0.448	0.107	0.087	0.113	0.126	
	0.015	0.005	0.019	0.014	0.005	0.014	0.062	
4.RAG	220	105	81	768	144	396	1188	2902
	0.076	0.036	0.028	0.265	0.050	0.136	0.409	0.149
	0.111	0.116	0.098	0.304	0.125	0.161	0.123	
	0.011	0.005	0.004	0.039	0.007	0.020	0.061	
5.TER	191	135	59	279	455	355	1190	2664
	0.072	0.051	0.022	0.105	0.171	0.133	0.447	0.136
	0.097	0.150	0.071	0.110	0.394	0.144	0.123	
	0.010	0.007	0.003	0.014	0.023	0.018	0.061	
6.NEG	236	101	81	412	168	595	1440	3033
	0.078	0.033	0.027	0.136	0.055	0.196	0.475	0.155
	0.119	0.112	0.098	0.163	0.145	0.241	0.149	
	0.012	0.005	0.004	0.021	0.009	0.030	0.074	
7.NEU	221	104	66	369	129	358	2386	3633
	0.061	0.029	0.018	0.102	0.036	0.099	0.657	0.186
	0.112	0.115	0.080	0.146	0.112	0.145	0.247	
	0.011	0.005	0.003	0.019	0.007	0.018	0.122	
Column Total	1976	902	829	2526	1155	2465	9667	19520
	0.101	0.046	0.042	0.129	0.059	0.126	0.495	

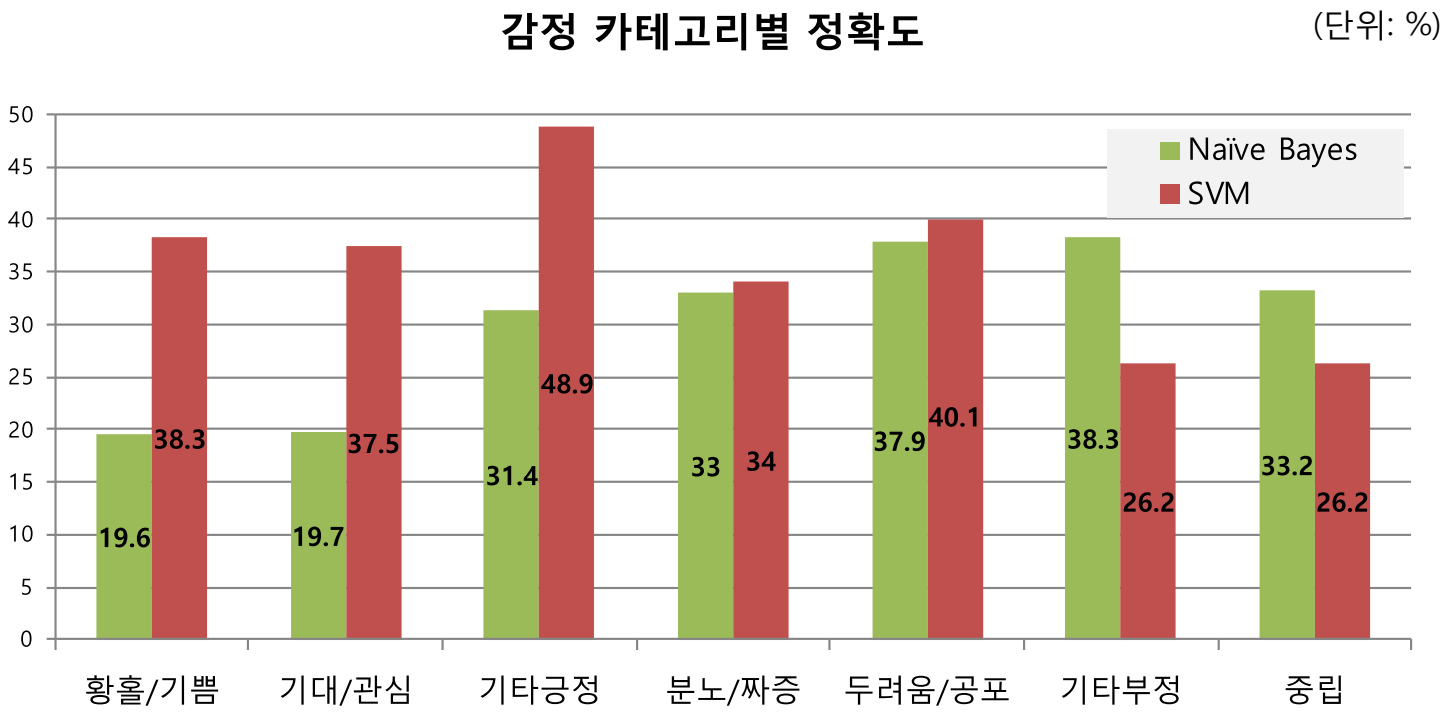
전체 데이터 : 7개 감성 class 분류 성과

Model Accuracy
- SVM1 (Linear Kernel) : 30.60%
- SVM2 (Polynomial Kernel) : 27.81%
- SVM3 (RBF Kernel) : 30.66%
- SVM4 (Sigmoid Kernel) : 27.47%



전체 데이터 : 7개 감성 class 분류 성과

Model Accuracy
- Naïve Bayes : 26.59%
- SVM(RBF Kernel) : 30.66%



1. 전체 데이터 : 감성 분류 정확도

(단위: %)

	Sentiment Class	
	7 Class	3 Class
nb_Model	26.59	47.69
svm_Model_1	30.60	52.57
svm_Model_2	27.81	51.38
svm_Model_3	30.66	54.94
svm_Model_4	27.47	50.38

2. 매체별 감성 분류 정확도

(단위: %)

Chanel	Model	Sentiment Class	
		7 Class	3 Class
Facebook	nb_Model	25.15	42.76
	svm_Model_1	41.25	59.71
	svm_Model_2	33.92	53.50
	svm_Model_3	42.22	58.31
	svm_Model_4	43.21	59.25
Twitter	nb_Model	20.36	43.88
	svm_Model_1	38.33	55.51
	svm_Model_2	30.80	50.14
	svm_Model_3	40.93	56.36
	svm_Model_4	40.30	55.53
Instagram	nb_Model	22.94	34.20
	svm_Model_1	37.61	63.14
	svm_Model_2	24.23	56.78
	svm_Model_3	36.96	61.96
	svm_Model_4	38.54	62.31
News comment	nb_Model	23.09	39.40
	svm_Model_1	35.18	59.81
	svm_Model_2	27.91	54.67
	svm_Model_3	34.58	60.10
	svm_Model_4	34.92	59.29

3. 현상별 감성 분류 정확도

(단위: %)

Chanel	Model	Sentiment Class	
		7 Class	3 Class
온도	nb_Model	23.56	48.57
	svm_Model_1	37.59	58.60
	svm_Model_2	30.94	52.65
	svm_Model_3	37.55	57.86
	svm_Model_4	37.92	58.52
강수	nb_Model	20.60	41.59
	svm_Model_1	35.44	58.47
	svm_Model_2	28.80	53.16
	svm_Model_3	36.90	58.12
	svm_Model_4	37.38	59.05
토지	nb_Model	24.62	37.41
	svm_Model_1	35.76	58.11
	svm_Model_2	26.03	53.88
	svm_Model_3	35.51	58.66
	svm_Model_4	34.71	58.11
해양	nb_Model	20.35	33.49
	svm_Model_1	37.04	59.38
	svm_Model_2	30.28	56.55
	svm_Model_3	37.20	58.60
	svm_Model_4	36.58	59.00

감성 분류 정확도

(단위: %)

		Sentiment Class	
		7 Class	3 Class
매체	Facebook	43.21	59.71
	Twitter	40.93	56.36
	Instagram	38.54	63.14
	News comment	35.18	60.10
현상	온도	37.92	58.60
	강수	37.38	59.05
	토지	35.76	58.66
	해양	37.20	59.38

Research Finding

1. 감성분류별 분석 결과
   - 3 class 분류가 7 class 보다 약 20~25% 정확도 높음
   - 3 class 분류 : Linear Kernel SVM 성능이 가장 좋음
   - 7 class 분류 : Sigmoid Kernel SVM 성능이 가장 좋음

3. 매체별 분석 결과
   - 7 class 분류 : 약 13% 정확도 향상
   - 3 class 분류 : 약 9% 정확도 향상
   - 인스타그램의 경우, 3class 분류 시 최고 정확도(63.14%) 나타남
   - 페이스북의 경우, 7class 분류 시 최고 정확도(43.21%) 나타남
   - 이모지를 주로 사용하는 인스타그램, 페이스북의 감성분류 성능이 높음

4. 현상별 분석 결과
   - 7 class 분류 : 약 5% 정확도 향상
   - 3 class 분류 : 약 7% 정확도 향상
   - 현상별 감성분류 성능 차이는 1~2%로 매우 낮음

향후 계획

- 임베딩과 딥러닝을 이용한 분류기 구축
  - CNN, RNN, LSTM, GRU 등의 딥러닝 기반 분류 알고리즘 구축
  - 분류 정확도 성능 평가를 통해 최종 분류모델 선정

	8월 초	9월 말
전처리	1) 이모지 한글로 변환 (😍 , 😱) 2) 이모티콘(특수문자) 전처리 (ㅠㅠ, ^^) 3) ID 삭제 (@lloook91) 4) 형태소 분석 4) **정규화** : 함축어, 신조어, 은어 등	
	5) Document Term Matrix(DTM) 생성 7) Sparse Terms 삭제	5) Token Embedding Matrix 생성 : token -> vector (단어 의미 부여)
구현	8) Naïve bayes 9) SVM	6) CNN 7) RNN 8) LSTM 9) GRU