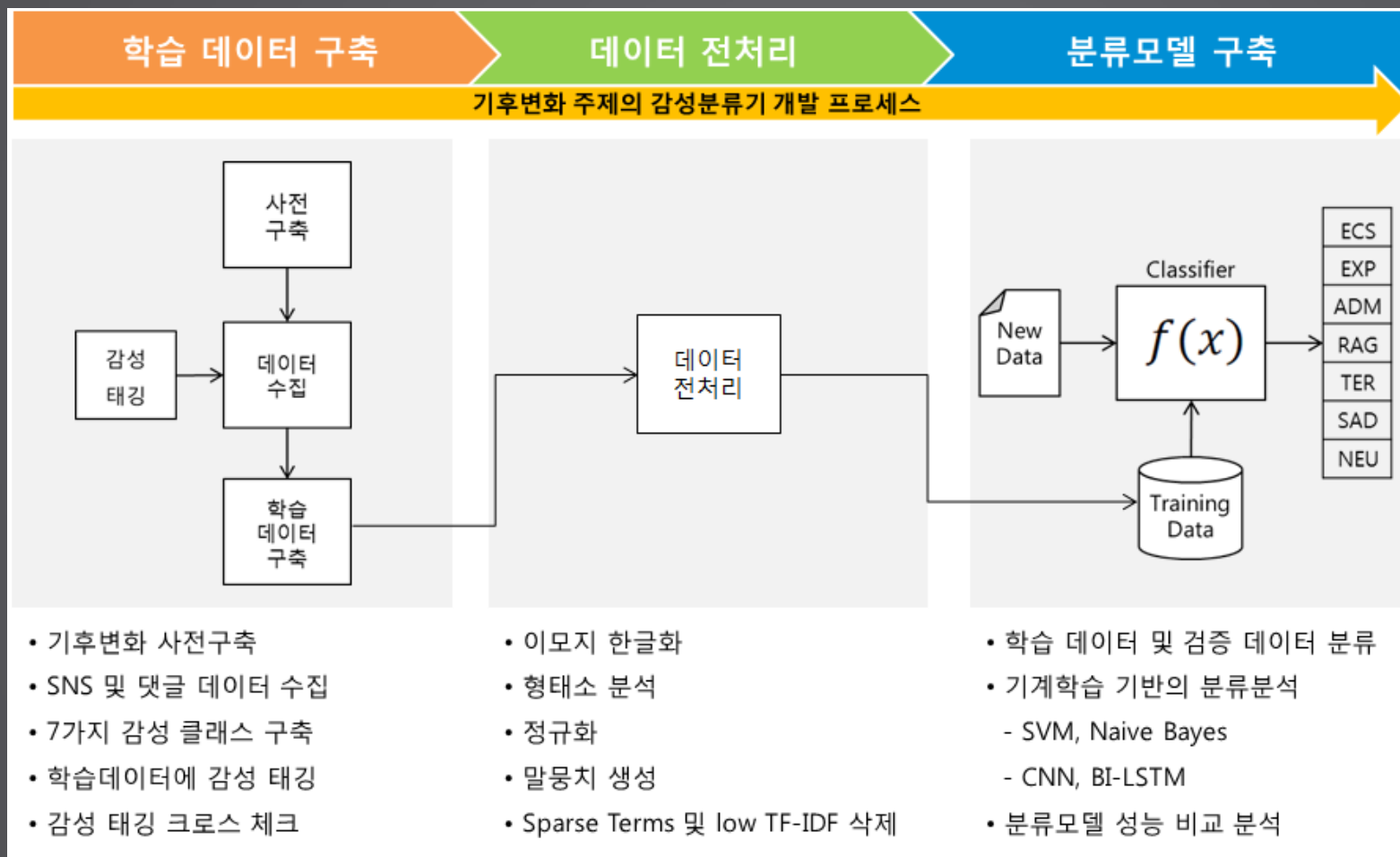


(3) 기계학습 기반 환경이슈 감성분류기 개발 : 기후변화 중심으로 [김도연]

- ◆ 연구 목적: 기후변화 주제의 SNS 및 뉴스 댓글 데이터 기반 감성분류 알고리즘 개발
- ◆ 연구 내용: 기후변화 사전 구축, 감성 분류 학습 데이터 구축, 감성분류 알고리즘 개발
 - **기후변화 사전** : 기후변화에 따른 현상을 4개의 범주(온도, 강수, 토지, 해양) 분류 후 구축
 - 환경관련 문서에 워드 임베딩 방법(LDA, Word2Vec) 적용 후보군 추출
 - 전문가(최희선, 명수정) 및 SNS 이용자 의견 반영
 - **감성분류 기준표** : 기후변화 현상에 자주 나타나는 7개 감성 클래스 구축
 - 7개 감성 카테고리 : 황홀/기쁨, 기대/관심, 감탄/존경, 분노/짜증, 두려움/공포, 슬픔/수심, 중립
 - **감성분류 학습데이터** : 5만 건 단문 데이터에 감성을 수작업으로 파악
 - 기후변화 사전 기준 5만건을 수집하여 7개의 감성 클래스 태깅
 - **감성분류 알고리즘** : 기계학습 기반 분류 알고리즘 구축
 - Naïve Bayes, SVM, CNN, Bi-directional LSTM
- ◆ 연구 성과 : 7개 감성 Category 대상 정확도 85.10% Bidirectional LSTM 알고리즘 구축
 - Accuracy : (True Positive + True Negative)/ALL sample

연구 범위 및 흐름도

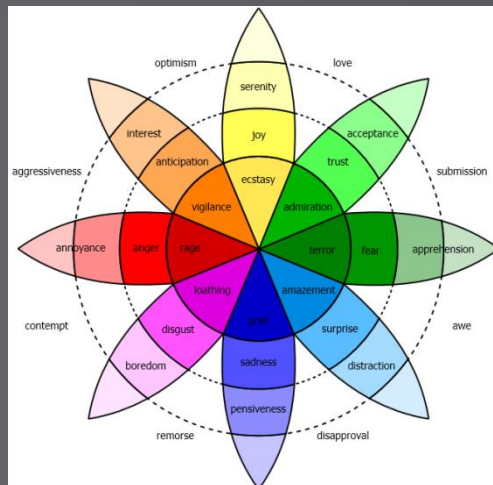


기후변화 사전 및 감성 클래스

기후변화에 따른 현상 사전

구분	순번	온도	강수	토지	해양
전문가	1	강추위	대설	가뭄	녹조
	2	결빙	산성비	사막화	라니냐
	3	무더위	우박	산불	쓰나미
	4	열대야	장마	산사태	엘니뇨
	5	열섬	적설	열대림파괴	적조
	6	열파	집중강우	지진	침수
	7	온난	집중호우	토지황폐화	파랑
	8	온실가스	폭설	화산폭발	풍랑
	9	이상고온	폭우	-	풍수해
	10	이상기온	홍수	-	해랑
	11	이상저온	황사비	-	해수면
	12	폭염	-	-	해일
	13	한파	-	-	-
	14	혹서	-	-	-
	15	혹한	-	-	-
비전문가	16	짙출	눈난리	갈라진땅	피물파도
	17	짙덥	눈쓰레기	메마른땅	큰파도
	18	쫄쫄	눈폭탄	산폭발	-
	19	쫄덥	물난리	찢어진땅	-
	20	넙덥	비폭탄	흔들리는땅	-
	21	넙출	홍비	-	-
	22	너무출	홍탕물비	-	-
	23	너무덥	-	-	-
	24	개출	-	-	-
	25	개덥	-	-	-

감성분류 기준표



감성 구분		감성 태그
긍정	황홀/기쁨	ECS
	기대/관심	EXP
	감탄/존경	ADM
부정	분노/짜증	RAG
	두려움/공포	TER
	슬픔/수심	SAD
중립		NEU

학습 데이터 구축 및 전처리

1. 기후변화에 따른 현상 사전 구축

구분	no.	기후변화에 따른 '현상' 키워드			
전문가	1	온도	강수	토지	해양
	2	강추위	대설	가뭄	녹조
	3	결빙	산성비	사막화	라니냐
	4	무더위	우박	산불	쓰나미
	5	열대야	장마	산사태	엘니뇨
	6	열성	적설	열대림파괴	적조
	7	열파	집중강우	지진	침수
	8	온난	집중호우	토지황폐화	파랑
	9	온실가스	폭설	화산폭발	홍수
	10	이상고온	폭우		풍수해
	11	이상기온	홍수		해양
	12	이상저온	황사비		해수면
	13	폭염			해일
	14	한파			
	15	혹서			
비전문가	16	정수	눈난리	갈라진땅	괴물파도
	17	정답	눈쓰레기	메마른땅	큰파도
	18	출출	눈폭탄	산폭발	
	19	출입	눈난리	뿔어진땅	
	20	넙넙	비폭탄	흔들리는땅	
	21	넙넙	홍비		
	22	너무습	홍탕홍비		
	23	너무습			
	24	개습			
	25	개습			

2. 감성분류 기준 구축

	감성 태그	
긍정	황홀/기쁨	ECS
	기대/관심	EXP
	감탄/존경	ADM
부정	분노/짜증	RAG
	두려움/공포	TER
	슬픔/수심	SAD
중립	NEU	

3. 학습데이터 구축

No	Chanel	Keyword	Keyword1	Content	cross check					
					Tag	tag1	tag2	tag3	tag4	final Tag
1	뉴스 댓글	온도	강추위	난 추울때 겨울 냄새가 날 좋아	ADM	ECS	ADM	ECS	ECS	ECS
2	뉴스 댓글	온도	무더위	밖에서 일하시는 이 세상에 아버지 어머니를 힘내세요!!!	ADM	ADM	ADM	ECS	ECS	ADM
3	뉴스 댓글	온도	무더위	더 뜨거워지면 좋겠다	EXP	EXP	EXP	EXP	EXP	EXP
4	뉴스 댓글	온도	강추위	동요에 찬바람 불어도 괜찮아요 가사있죠? 이 날씨에 괜찮은 사람 없을까...	RAG	RAG	RAG	TER	TER	RAG
5	뉴스 댓글	온도	무더위	하루하루가 더위때문에 날 힘들네요..내일은 얼마나 또더울까..이런생각에..	TER	TER	TER	TER	TER	TER
6	뉴스 댓글	온도	너무습	의정부 -13 너무 좋다..ㅠ ㅠ	SAD	SAD	SAD	SAD	SAD	SAD
8	뉴스 댓글	온도	강추위	오늘 모스크바 -11 대관령 -21.7	NEU	NEU	NEU	NEU	NEU	NEU
9	Facebook	온도	열대야	사워하고먹는아이스아메리카노가 진리임 열대야극복완료 단순해ㅋㅋ	ECS	ECS	NEU	ECS	ADM	ECS
10	Facebook	온도	무더위	밖에 내다 놓은 화조들이 무더위를 견디다니..러려..튼튼한것들	ADM	ADM	ADM	NEU	NEU	ADM
11	Facebook	온도	열대야	열대야가 이번주가 끝이래요!!! 밖으로 놀러가자아~~~~	EXP	EXP	ADM	EXP	EXP	EXP
12	Facebook	온도	결빙	이젠 진심 눈이 싫어 지네요	RAG	RAG	TER	TER	RAG	RAG
13	Facebook	온도	결빙	역대급추위..... = = = = 을 첫 영상강 결빙현상 발생	TER	TER	TER	TER	TER	TER
14	Facebook	온도	무더위	음? 무더위 본격적 시작이었는데 .. 그리고 10월까지 없다는데 실화냐 ㅠ	SAD	SAD	TER	TER	TER	TER

전처리 단계

전처리 내용

- 1) 이모지 한글로 변환
- 2) 이모티콘(특수문자) 전처리
- 3) 형태소 분석
- 4) ID 삭제
- 5) 정규화: 합축어, 신조어, 은어

- SNS 특성을 반영한 전처리 단계
- 형태소 분석기 : 은전한뿔-Mecab
- 이모지 전처리: 약 1,200개 이모지 한글로 변환
- 예) 😍 😱

- 6) DTM 생성
- 7) 말뭉치(Corpus) 생성
- 8) Sparse Terms 삭제
- 9) Low TF-IDF 삭제

- DTM(Document Term Matrix)

	Term1	Term2	...	TermM
Doc1	2	1	...	0
Doc2	0	4	...	2
...				...
DocN	3	1	...	1

SVM Classifier

- 커널(Kernel) 트릭을 이용한 비선형 데이터 분류
- 커널 종류 및 파라미터
 - 1) 선형 커널(Linear Kernel) : Cost, Gamma
 - 2) RBF 커널(Radial Basis Function Kernel) : Cost, Gamma
 - 3) 시그모이드 커널(Sigmoid Kernel) : Cost, Gamma, Coefficient
 - 4) 다항식 커널(Polynomial Kernel) : Cost, Gamma, Coefficient, Degree
- SVM Architecture

Parameter	Model			
	SVM_Model_1	SVM_Model_2	SVM_Model_3	SVM_Model_4
Kernel	Linear	RBF	Sigmoid	Polynomial
Cost	1	0.8	1	0.9
Gamma	0.0048	0.0040	0.0047	0.00047
Coefficient	-	-	0.001	0.001
Degree	-	-	-	3

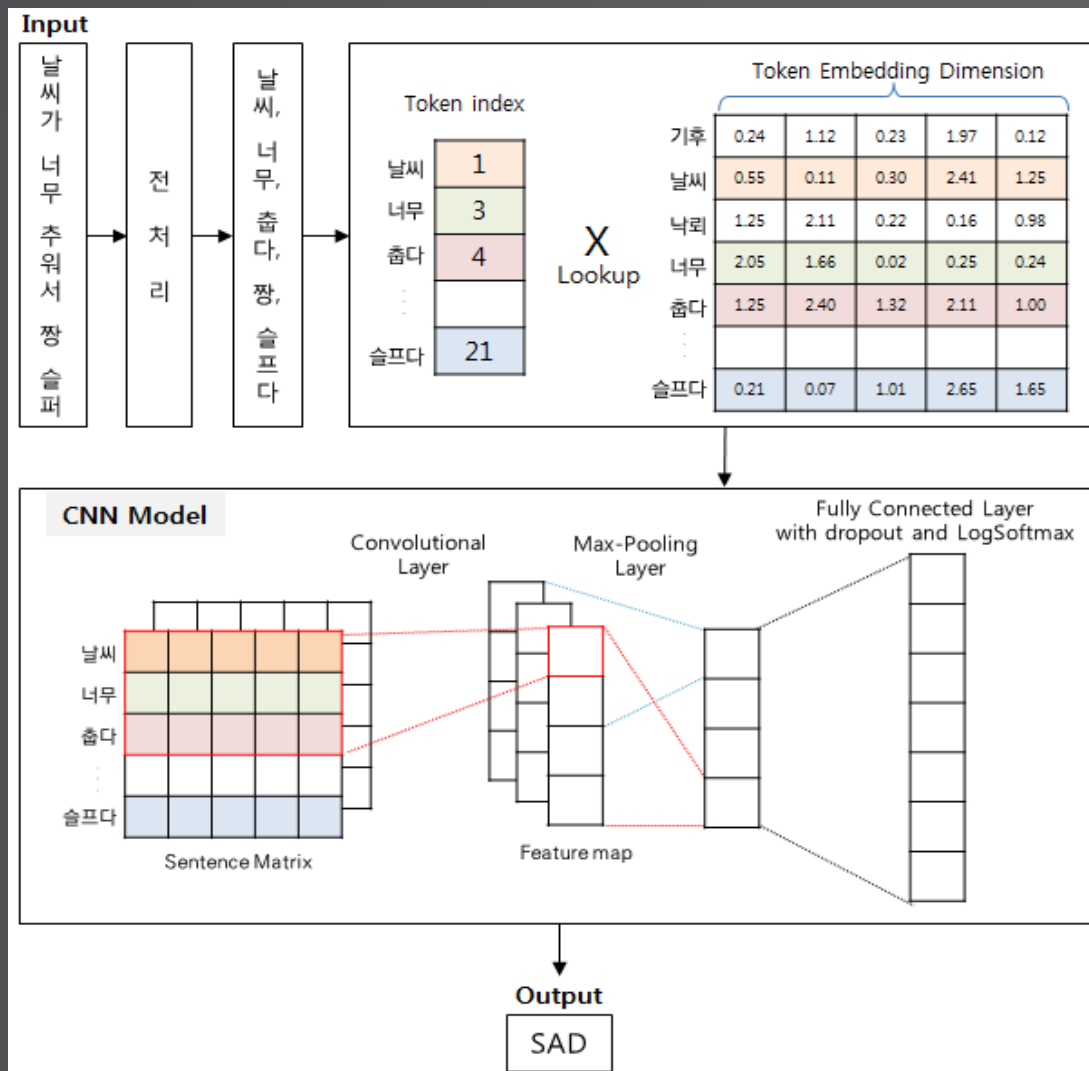
$$\text{Linear Kernel} : K(x_n, x_i) = (x_n, x_i)$$

$$\text{RBF Kernel} : K(x_n, x_i) = \exp(-\gamma \|x_n - x_i\|^2 + C)$$

$$\text{Sigmoid Kernel} : K(x_n, x_i) = \tanh(\gamma(x_n, x_i) + r)$$

$$\text{Polynomial Kernel} : K(x_n, x_i) = (\gamma(x_n, x_i) + r)^d$$

CNN Classifier



CNN Classifier(

Embedding: dimension(13360, 128)

CNN-3-100 : Con2d(1, 100, kernel_size=(3, 128), stride=(1,1))

Activation : ReLU()

CNN-4-100 : Con2d(1, 100, kernel_size=(4, 128), stride=(1,1))

Activation : ReLU()

CNN-5-100 : Con2d(1, 100, kernel_size=(5, 128), stride=(1,1))

Activation : ReLU()

Generator : Linear(in_features=300, out_features=7)

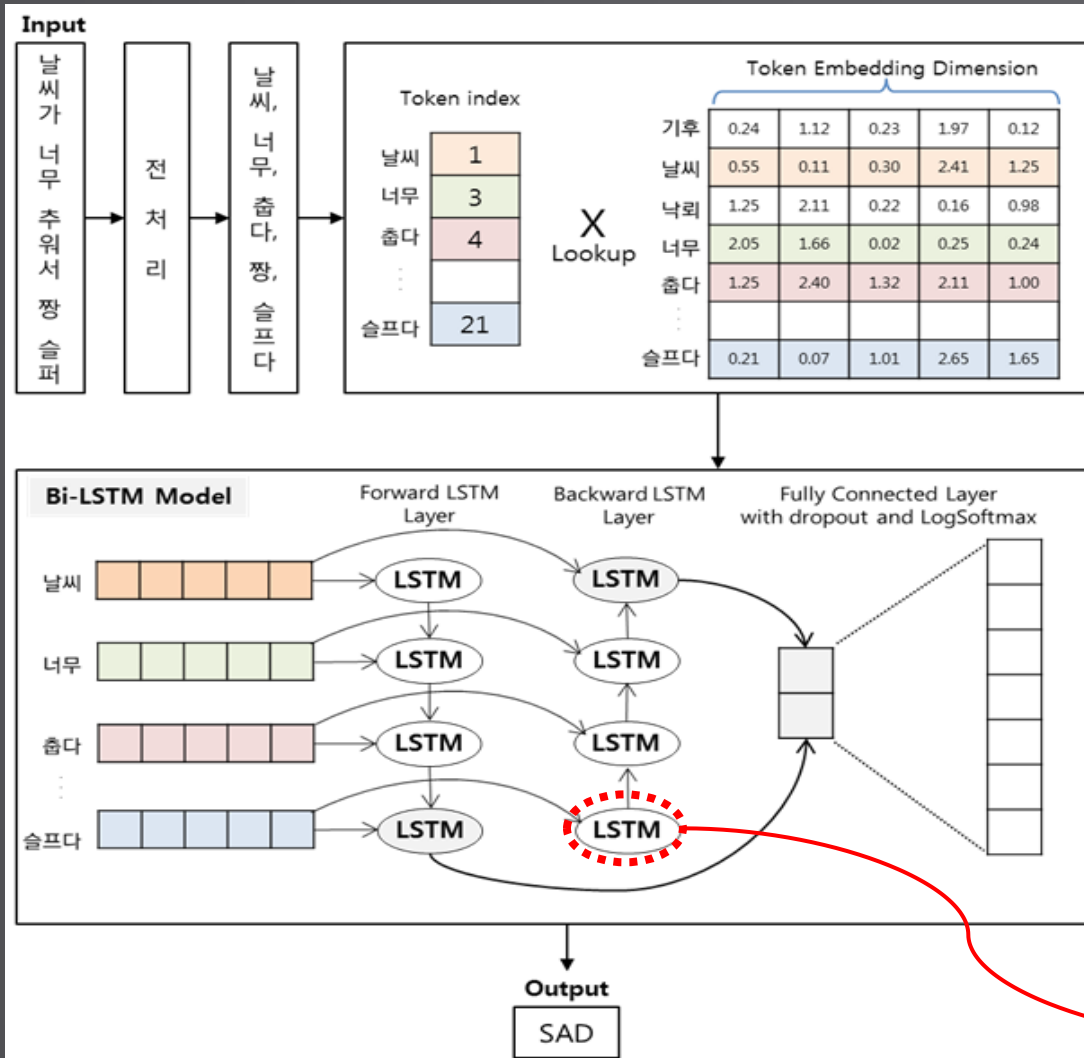
Pooling: Max-Pooling

Dropout : Dropout(p=0.3)

Activation: LogSoftmax()

)

Bidirectional LSTM Classifier



Bi-LSTM Classifier(

Embedding : dimension(13360, 128)

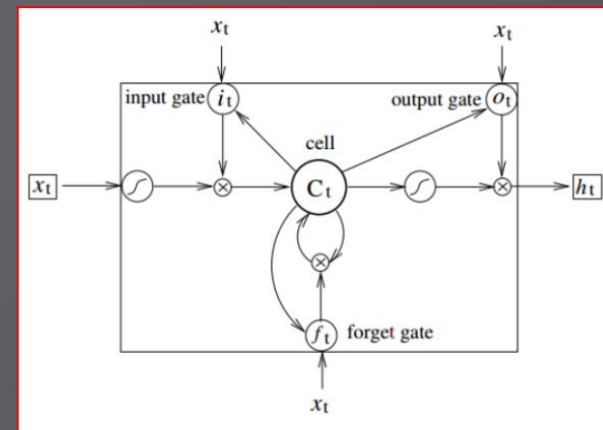
RNN: LSTM(num_layers = 4, bidirectional = True)

Generator : Linear(in_features=512, out_features=7)

Dropout : Dropout(p=0.3)

Activation: LogSoftmax()

)

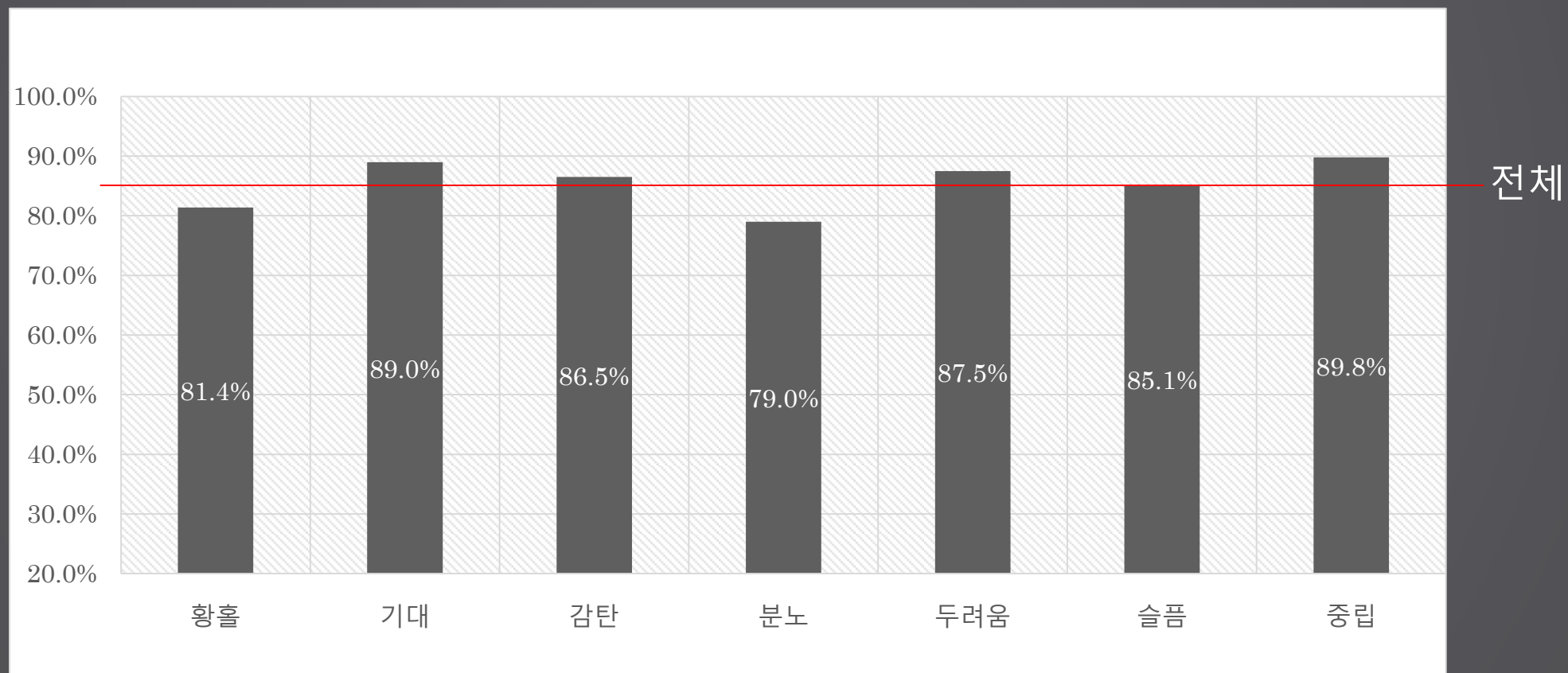


감성 분류 정확도(Accuracy) 비교

	Model	7개 감성 분류(%)	3개 감성 분류(%)
Machine Learning	SVM_Model_1	77.96	87.25
	SVM_Model_2	68.71	76.25
	SVM_Model_3	76.85	86.31
	SVM_Model_4	75.04	84.08
	NB_Model	50.57	65.73
Deep Learning	CNN_Model	81.95	90.65
	Bi-LSTM_Model	85.10	92.95

7 개 감정 class별 분류 성과 : Bi-LSTM 85.10%

감성 카테고리별 정확도



매체별 분류 정확도: Bi-LSTM 85.10%

매체별 정확도

