

# Fake News Classifier

새싹 도봉캠퍼스 김현서

	title	text	subject	date	fake
0	France invites U.S. to Dec. 13 summit on boost...	ACCRA (Reuters) - French President Emmanuel Ma...	worldnews	November 30, 2017	0
1	Trump keeps politics on his Thanksgiving menu	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - U...	politicsNews	November 24, 2016	0
2	Ireland's Fianna Fail party says will be elect...	DUBLIN (Reuters) - Ireland s second-largest pa...	worldnews	November 24, 2017	0
3	REPUBLICAN LEADER Implies He May Not Seat Roy ...	The New York Times just confirmed what we ve a...	politics	Nov 12, 2017	1
4	U.S. Senate approves Trump pick as top Fed reg...	WASHINGTON (Reuters) - The U.S. Senate on Thur...	politicsNews	October 5, 2017	0
5	SNARKY HILLARY AIDE EMBARRASSED BY Inconvenien...	There never seems to be a shortage of smug, kn...	politics	Feb 23, 2016	1
6	Obama Can't Stop Winning As New Jobs Report S...	Today, I would hate to be a Republican strateg...	News	January 8, 2016	1
7	While Her Department Gets Gutted, Billionaire...	Donald Trump and his administration seem to be...	News	April 10, 2017	1
8	(VIDEO) FORMER DEFENSE SECRETARY EXPRESSES CON...		politics	May 18, 2015	1
9	Story Of Severely Disabled Man And His Mom Sh...	By now, everyone has seen the monstrosity that...	News	June 24, 2017	1
10	Behind Democrats' email leak, U.S. experts see...	WASHINGTON (Reuters) - If the Russian governme...	politicsNews	July 25, 2016	0
11	Czechs vote for new parliament, wealthy busine...	PRAGUE (Reuters) - Czechs voted on Friday in a...	worldnews	October 19, 2017	0
12	Calm Down! Democrats Are Going To Stomp Trump...	After Tuesday s semi-coronation of Donald Trum...	News	May 4, 2016	1
13	Obama says expects Taliban to continue strateg...	ISE-SHIMA, Japan (Reuters) - U.S. President Ba...	politicsNews	May 26, 2016	0
14	Hillary Clinton Survives Another FBI Pantomime...	Clare, Bernis, Free Thought ProjectIn no surp...	US_News	November 7, 2016	1
15	Trump National Security Adviser Seriously Vio...	On the one hand, the most controversial figures in the u...	News	December 14, 2016	1
16	Barnier reassures Ireland EU stands with Dubli...	BRUSSELS (Reuters) - European Union Brexit neg...	worldnews	November 24, 2017	0
17	Dutch join Austria's lawsuit against German ro...	AMSTERDAM (Reuters) - The Dutch government sai...	worldnews	December 13, 2017	0
18	FORBES LIST OF "The World's Most Powerful Peop...	Hey Michelle Guess what? For the first time in...	left-news	Dec 17, 2016	1
19	THE VIEW WOMEN Go Off The Rails: Trump 'has to...		politics	Dec 12, 2016	1

DATA

X: text / y: fake로 지정하고 진행

# 전처리

1. 공백 텍스트 제거
2. 모든 알파벳을 소문자로 변환
3. 특수 기호 제거
4. 여러 개의 공백을 한 칸으로 통일
5. 불용어 제거

# 자연어 처리

1. Tokenizer를 사용해 텍스트 데이터 sequence
  - ✓ 인덱싱된 단어의 개수가 너무 많기 때문에 중요도 순으로 정렬 후 제거
2. 최대 문장의 개수에 맞춰 padding
  - ✓ train 데이터 셋과 test 데이터 셋의 크기를 같게 맞춰준다

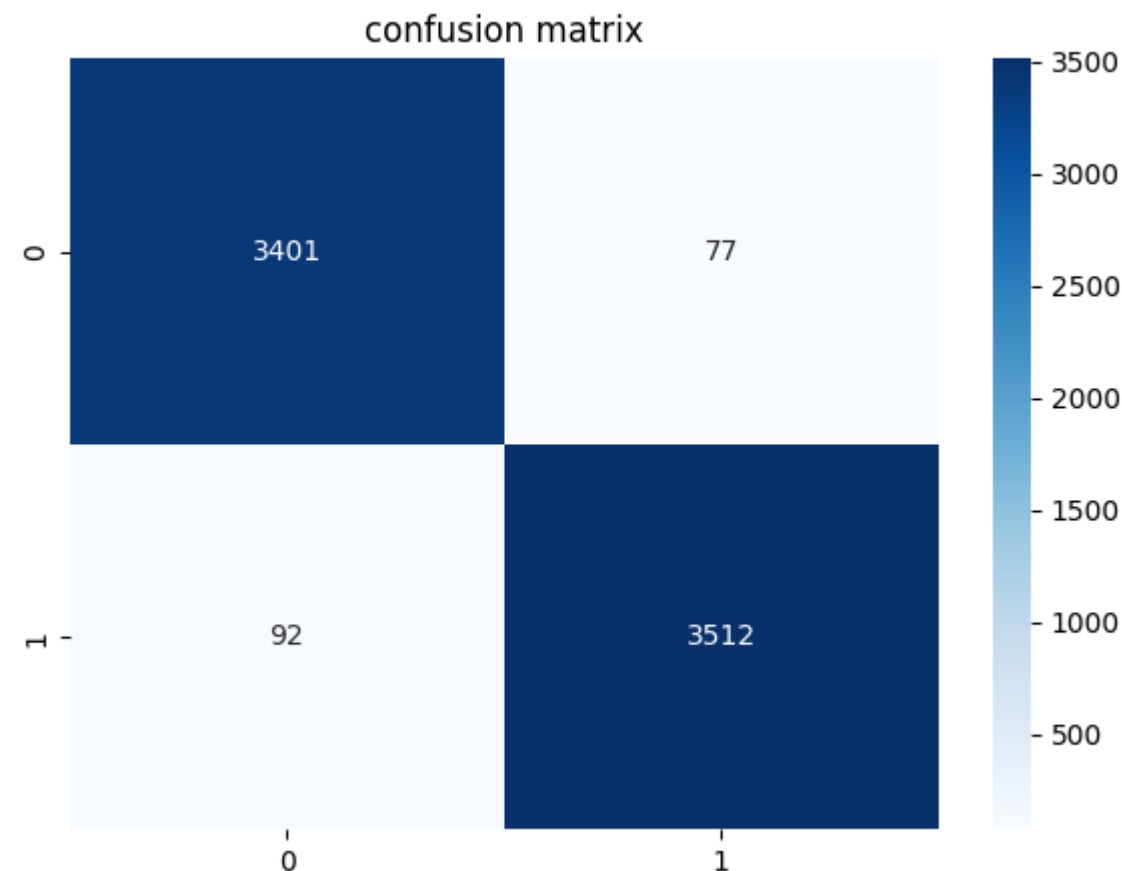
# Random Forest Classifier

```
rf = RandomForestClassifier(random_state=42, n_estimators=100)
rf.fit(X_train_padding, y_train)
```

정확도	정밀도
0.976	0.979
재현율	F1
0.974	0.977

## 개선 방안

- ✓ GridSearch를 통해 하이퍼파라미터 조절
- ✓ 모델에 더 특화된 데이터로 전처리



# SimpleRNN

```
edim = 10
hunits = 32

rnn = Sequential()

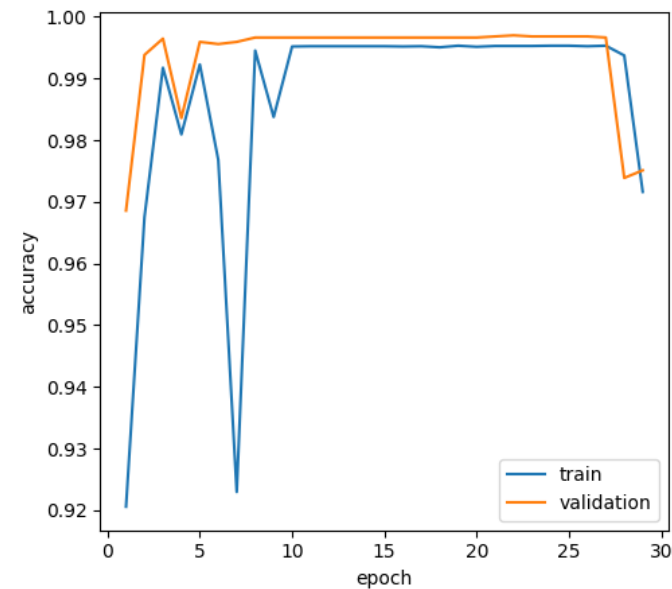
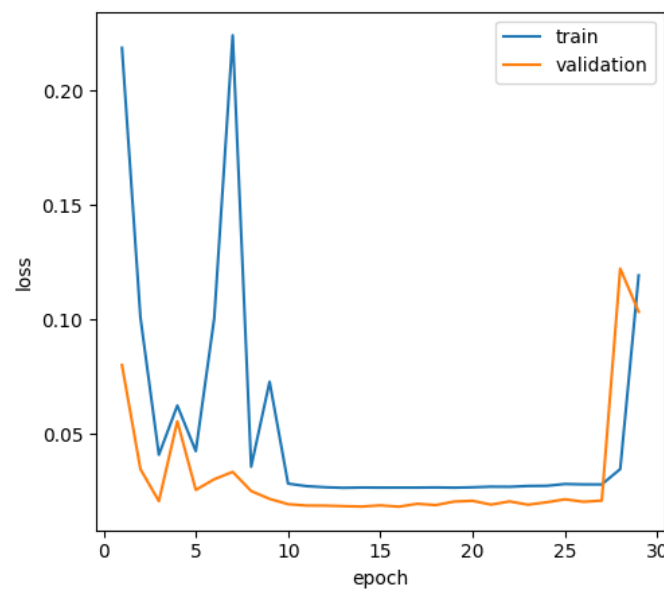
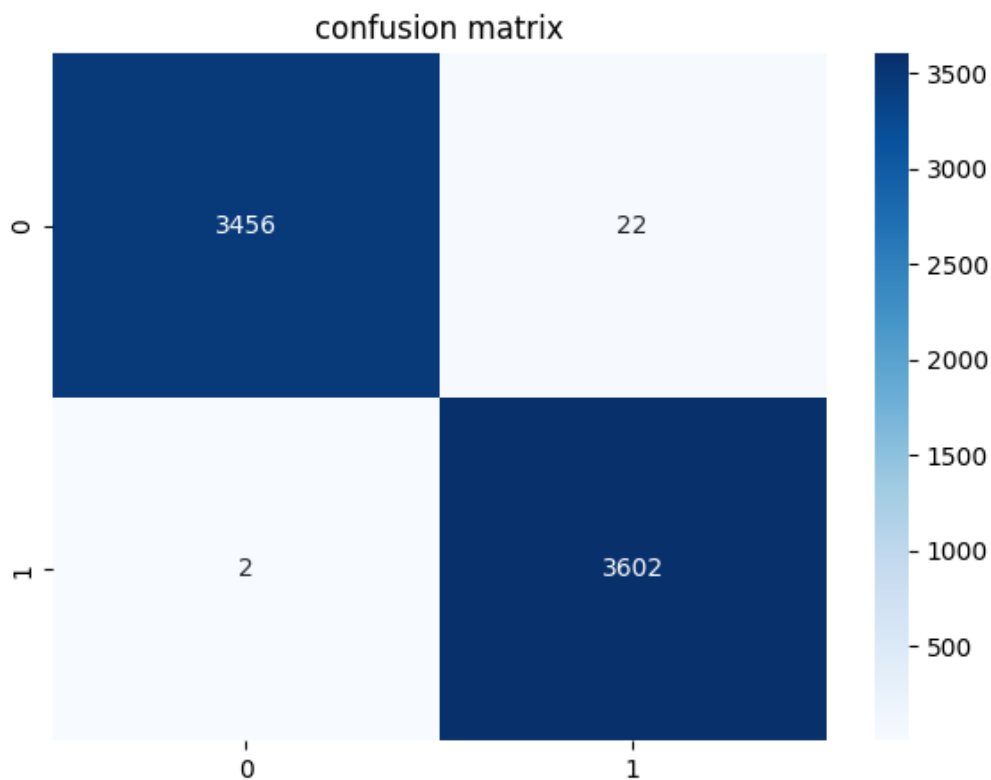
rnn.add(Embedding(vocab_size_mini, edim))
rnn.add(SimpleRNN(hunits))
rnn.add(Dense(1, activation='sigmoid'))

rnn.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
es = EarlyStopping(monitor='accuracy', patience=10, restore_best_weights=True)

rnnhist = rnn.fit(X_train_padding, y_train, epochs=50, validation_split=0.2, callbacks=[es])
```

정확도
0.9927

# SimpleRNN



## 개선 방안

- ✓ batch\_size 확대 (많은 데이터 학습을 위해)
- ✓ 축소된 단어 집합 다시 확대 후 탐색