

## 연습문제 – TF-IDF 연습 (Q12.1)

다음 문서 3개에 대하여 단어 빈도를 표시한 것이다. tf-idf 값을 구하여라.

	Doc1	Doc2	Doc3
car	27	4	24
auto	3	33	0
insurance	0	33	29
best	14	0	17

term	$df_t$	$idf_t$
car	18,165	1.65
auto	6723	2.08
insurance	19,241	1.62
best	25,235	1.5

Solution

	Doc1	Doc2	Doc3
car	44.55	6.6	39.6
Auto	6.24	68.64	0
Insurance	0	53.46	46.98

## 연습문제 – 영화리뷰 데이터 감성 분석을 수행하여 결과를 나타내어라 결과를 해석하여라 (Q12.2)

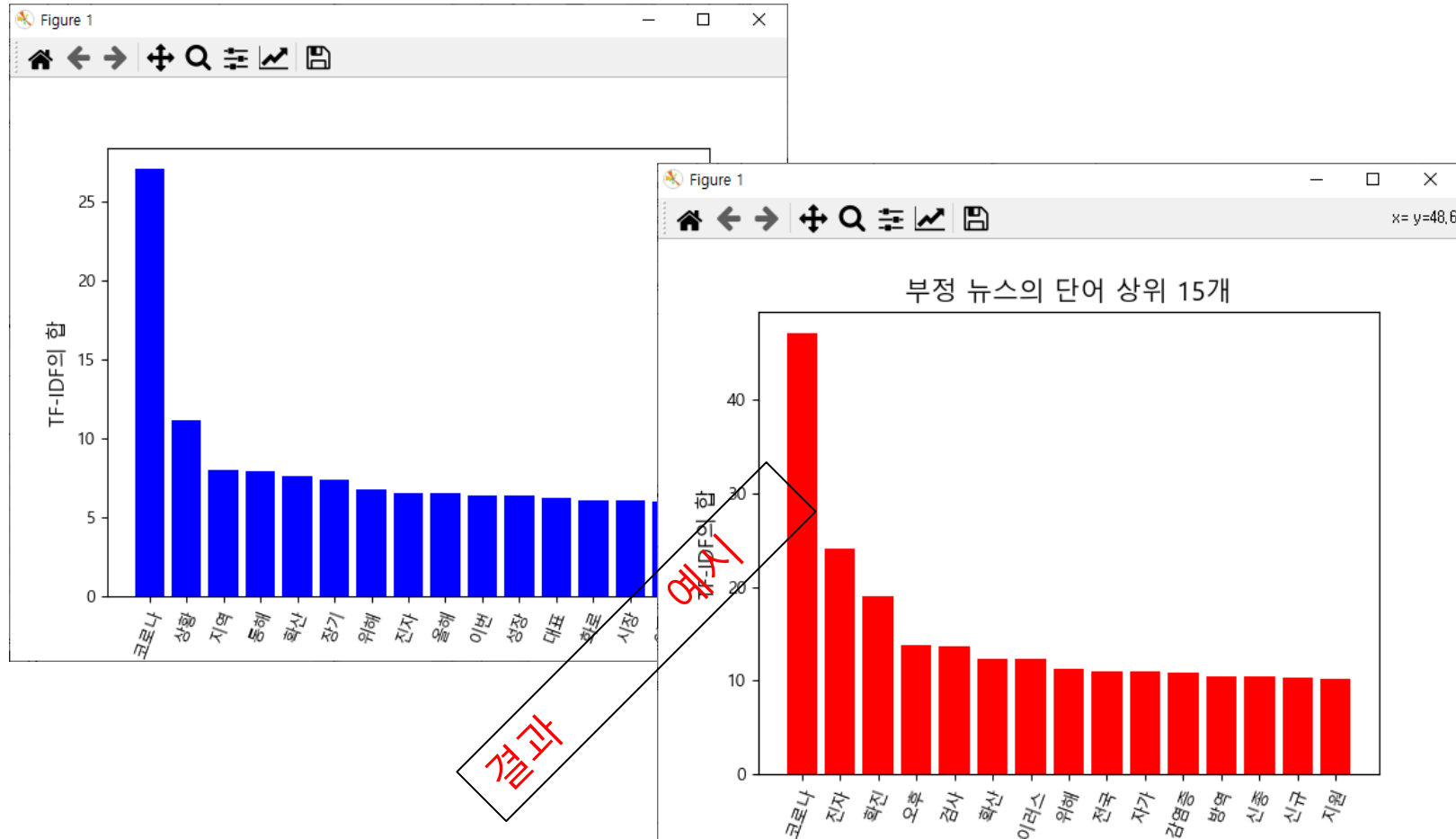
(1) 다음 감성에 대한 결과를 구해보자 – “크게 웃었네요”

(2) 본인의 리뷰를 넣어서 테스트 해보자

In [26]:	<code>st = input('감성 분석할 문장 입력 &gt;&gt; ')</code>
Out[26]:	감성 분석할 문장 입력 >> 웃자 ^o^ 오늘은 좋은 날이 될 것 같은 예감100%! ^^*
In [27]:	<code>#0) 입력 텍스트에 대한 전처리 수행 st = re.compile(r'[ㄱ-ㅣ가-힣]+').findall(st) print(st) st = [" ".join(st)] print(st)</code>
Out[27]:	'웃자', '오늘은', '좋은', '날이', '될', '것', '같은', '예감'] ['웃자 오늘은 좋은 날이 될 것 같은 예감']
In [28]:	<code>#1) 입력 텍스트의 피쳐 벡터화 st_tfidf = tfidf.transform(st) #2) 최적 감성 분석 모델에 적용하여 감성 분석 평가 st_predict = SA_lr_best.predict(st_tfidf)</code>
In [29]:	<code>#3) 예측값 출력하기 if(st_predict == 0):     print(st, "-&gt;&gt; 부정 감성") else :     print(st, "-&gt;&gt; 긍정 감성")</code>
Out[29]:	['웃자 오늘은 좋은 날이 될 것 같은 예감'] ->> 긍정 감성

(Q12.3, Q12.4는 안풀어도 감점없음)

연습문제 - 2부. 감성 분석 수행을 수행하여 결과를 나타내어라 결과를 해석하여라 (Q12.3)



## 연습문제 – 3부. 토픽모델링 : LDA 기반 토픽 모델링을 수행하여 결과를 나타내어라 결과를 해석하여라 (Q12.4)

