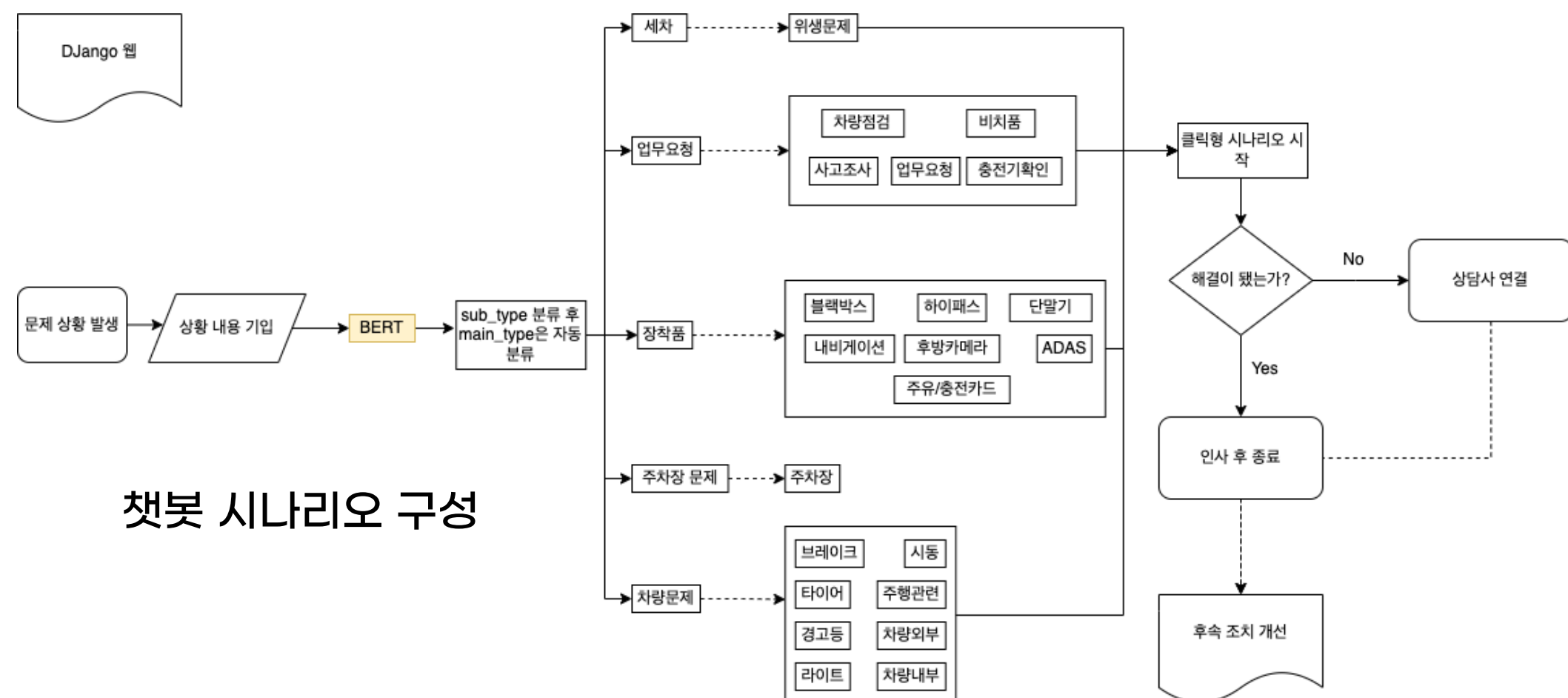




문제인식 및 주제 선정

- ✓ 메모 데이터에서, 고객 문의가 들어왔을 때 직접 카테고리를 분류하고 있음을 발견.
- ✓ 고객 문의가 들어오면 기본 카테고리 분류, 기본적인 시나리오 자동화 하는 것으로 주제 선정.
- ✓ 이로 인한 상담사 업무 피로도 감소, 불필요한 시간, 비용, 감정소비 감소 효과 기대.



챗봇 시나리오 구성



실험

- ✓ RNN 계열의 Sequential 모델로 oversampling, 형태소 분석 비교
- ✓ Transformer모델, 분류 decoder로 oversampling, 형태소 분석 비교
- ✓ Bert모델로 다양한 pretrained, oversampling 비교
- ✓ 상황내용 인지 후 BERT 모델을 통한 21가지 카테고리 중 분류
- ✓ 카테고리에 맞는 클릭형 시나리오 도출

Model Experiment

Sequential (RNN)

	Model	train_acc	train_loss	val_acc	val_loss	test_acc	test_loss
0	Baseline	0.9007	0.3453	0.7683	1.1689	0.7263	1.3240
1	Word2Vec	0.8677	0.4126	0.6500	1.6168	0.6091	1.8558
2	OverSampling	0.9659	0.1127	0.9400	0.2860	0.9272	0.2818
3	OverSampling+Word2Vec	0.8783	0.3591	0.8400	0.6284	0.8402	0.6815

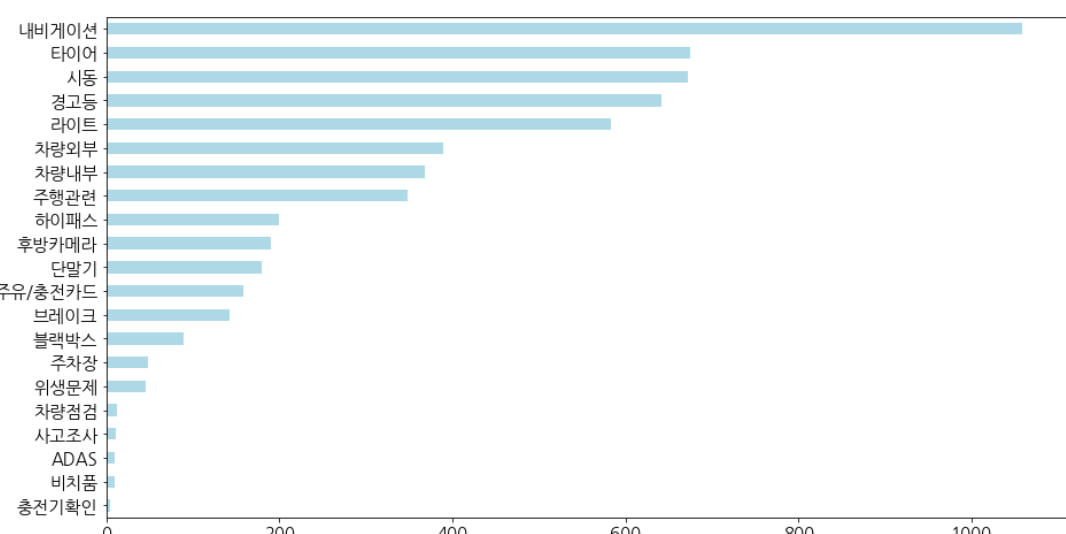
Transformer

	Model	train_acc	train_loss
0	Baseline	0.0800	0.0069
1	Okt	0.0827	0.0091
2	Okt+Word2Vec	0.0827	0.0087
3	Stopwords	0.1318	0.0111
4	단순 복제	0.1887	0.0074
5	OverSampling	0.1921	0.0053

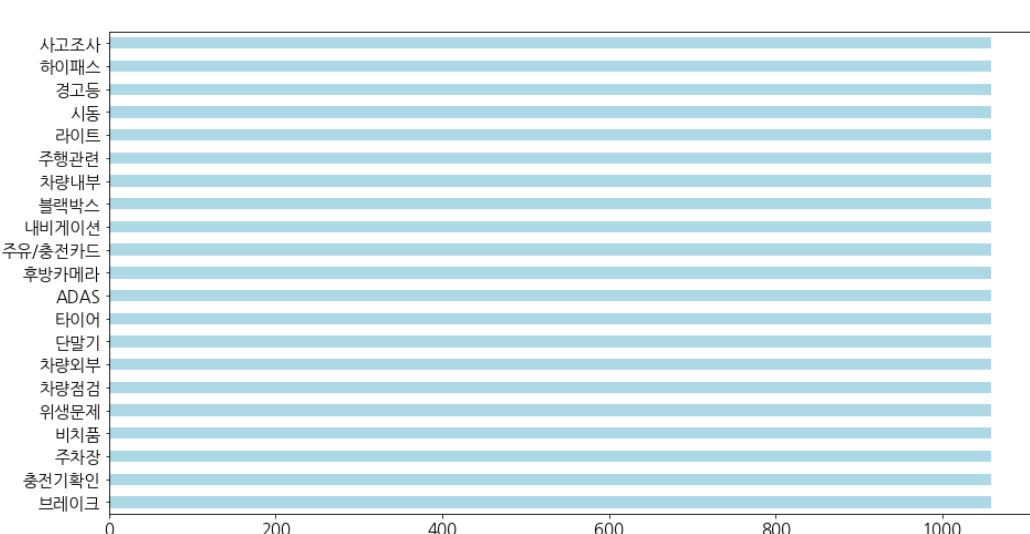
BERT

	Model	train_acc	train_loss	test_acc	test_loss
0	bert-base	0.0818	3.0445	0.0822	3.0445
1	bert-uncased	0.0167	3.0445	0.0017	3.0445
2	bert-base-multilingual-cased	0.1857	2.9173	0.1789	3.1299
3	TFBertModel	0.8869	0.4181	0.7586	1.0214
4	TFBertModel+OverSampling	0.9511	0.1989	0.9370	0.2668

Class Imbalance



OverSampling



결과

‘SubwordTextEncoder’를 이용한 단어장 생성

Bert 모델의 score 및 분류 결과

		precision	recall	f1-score	support	
{0: '<PAD>', 1: '<BOS>', 2: '<UNK>', 3: '<UNUSED>', 4: '인입', 5: '이_', 6: '경_', 7: '차량_', 8: '고등_', 9: '문의', 10: '가_', 11: '으로_', 12: '점등_', 13: '안_', 14: '타이어_',	0	0.92	0.84	0.88	212	
	1	0.92	0.81	0.86	212	
	2	0.79	0.84	0.82	211	
	3	0.90	0.90	0.90	212	
	4	0.83	0.75	0.79	212	
	5	0.93	0.91	0.92	211	
	6	0.88	0.90	0.89	212	
	7	0.90	0.92	0.91	212	
	8	1.00	1.00	1.00	212	
	9	0.90	0.91	0.90	211	
	10	0.98	0.99	0.98	212	
	11	0.95	0.92	0.94	212	
	12	0.95	0.99	0.97	211	
	13	1.00	1.00	1.00	211	
	14	0.98	1.00	0.99	212	
		15	0.95	1.00	0.97	212
		16	0.95	1.00	0.97	212
		17	0.97	1.00	0.98	212
		18	0.99	1.00	1.00	212
		19	1.00	1.00	1.00	211
	20	1.00	1.00	1.00	211	
accuracy				0.94	4444	
macro avg				0.94	4444	
weighted avg				0.94	4444	

장고 챗봇 구현