

Turkish Sentiment Analysis Data - beyazperde.com

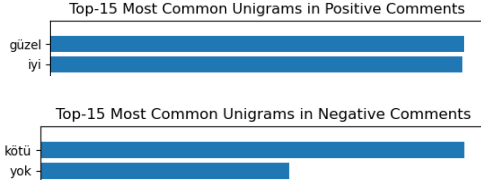
Test ve Train olarak ayrı iki dosya olarak paylaşılmış.

Train Dosyası:

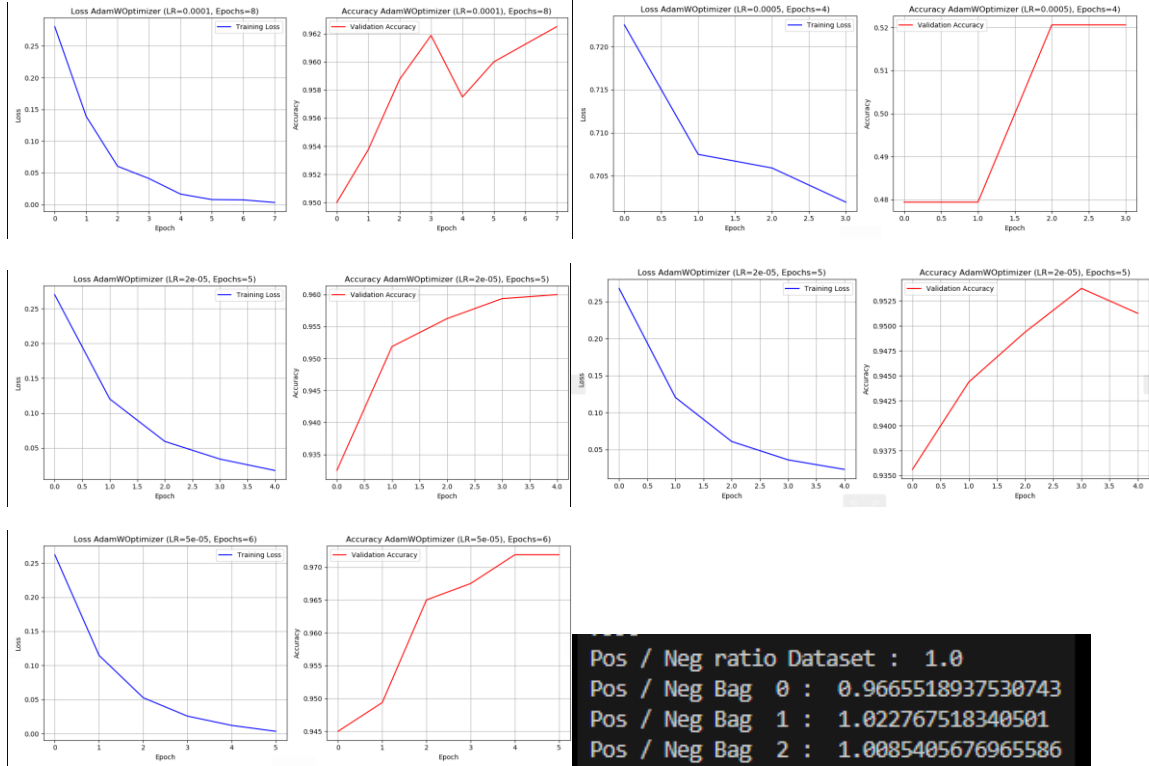
Label	Data Sayısı
1 (Pozitif)	3998
0 (Negatif)	3998

Test Dosyası:

Label	Data Sayısı
1 (Pozitif) - 0 (Negatif)	2666



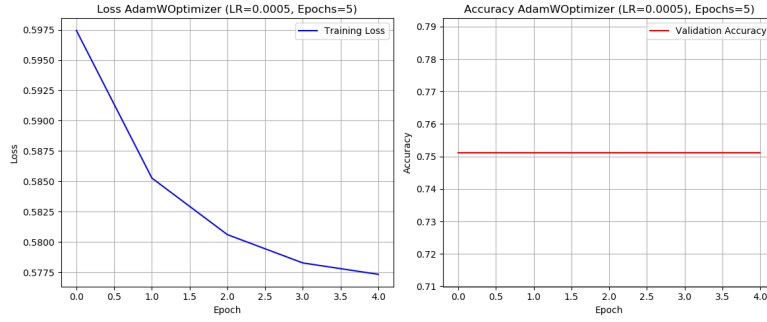
[1] <https://www.kaggle.com/code/ozcan15/turkish-sentiment-analysis-with-berturk-93-acc#1>-
Exploratory-Data-Analysis

Fine Tuning (Big Bert)

Training süresince bir epoch süresi: 0:02:15 dk sürmüştür.

	Bag0	Bag1	Bag2	Bag3	Bag4	Ensemble
--	------	------	------	------	------	----------

Accuracy(f1)	%92,52	%50	%92,498	%92,76	%92,198	%93,06
--------------	--------	-----	---------	--------	---------	--------



Bag 98 için çıktı(Ensemble'a dahil edilmemiştir).

Hiper Parametreler:

Epoch sayısı: 5

Learning rate: 0.0005

Optimizer: Adam Optimizer

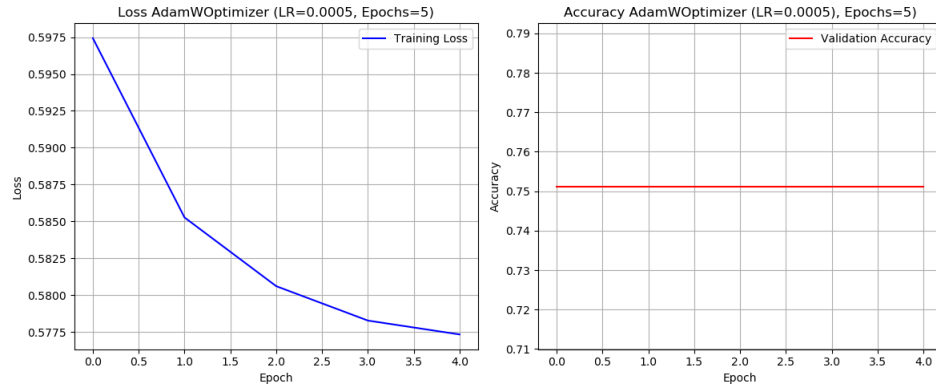
Batch size : 32

Training epoch took: 0:02:16

Accuracy: 0.75

Average training loss: 0.58

Pos / Neg Bag 3 : **0.344091443940158**



Bag 99 için çıktı. (Ensemble'a dahil edilmemiştir).

Hiper Parametreler:

Epoch sayısı: 5

Learning rate: 0.0005

Optimizer: Adam Optimizer

Alperen Kösem

Batch size : 32

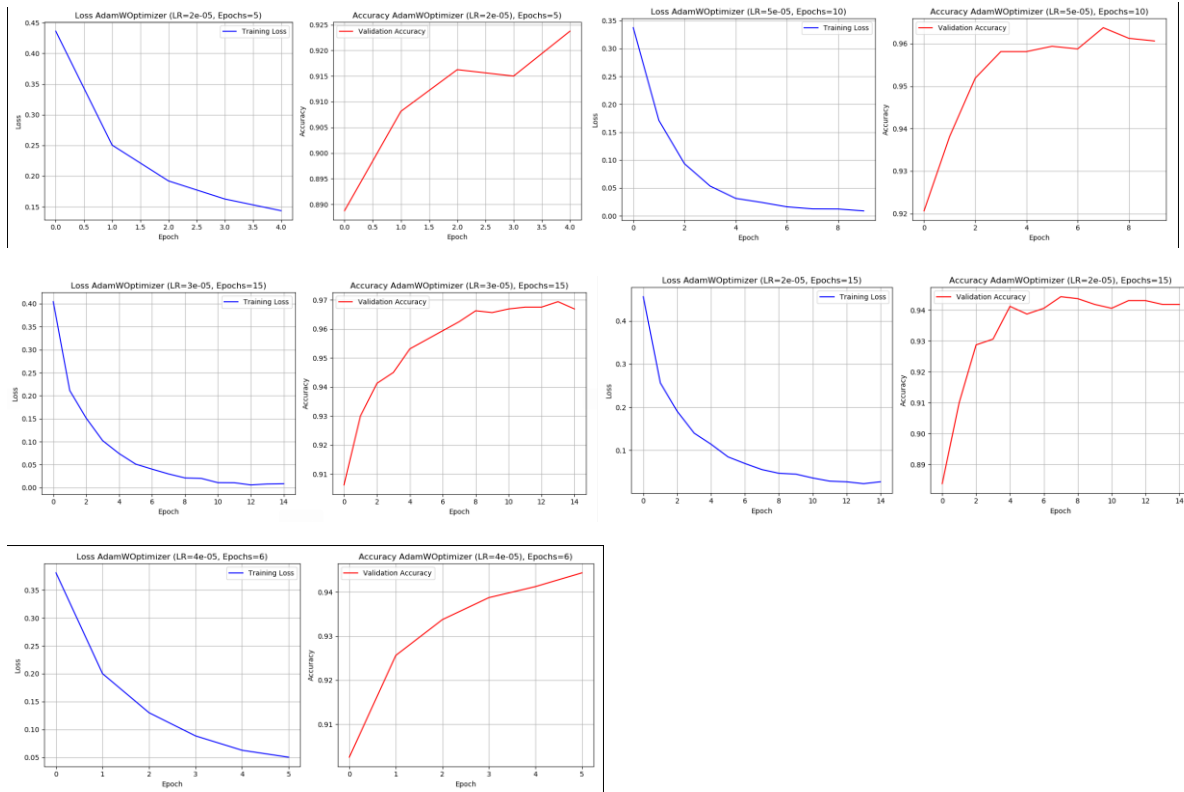
Training epoch took: 0:02:16

Accuracy: 0.77

Average training loss: 0.57

Pos / Neg Bag 4 : **3.00400600901352**

Fine Tuning (Small Bert)



training süresince bir epoch süresi: 0:00:18 sn sürmüştür

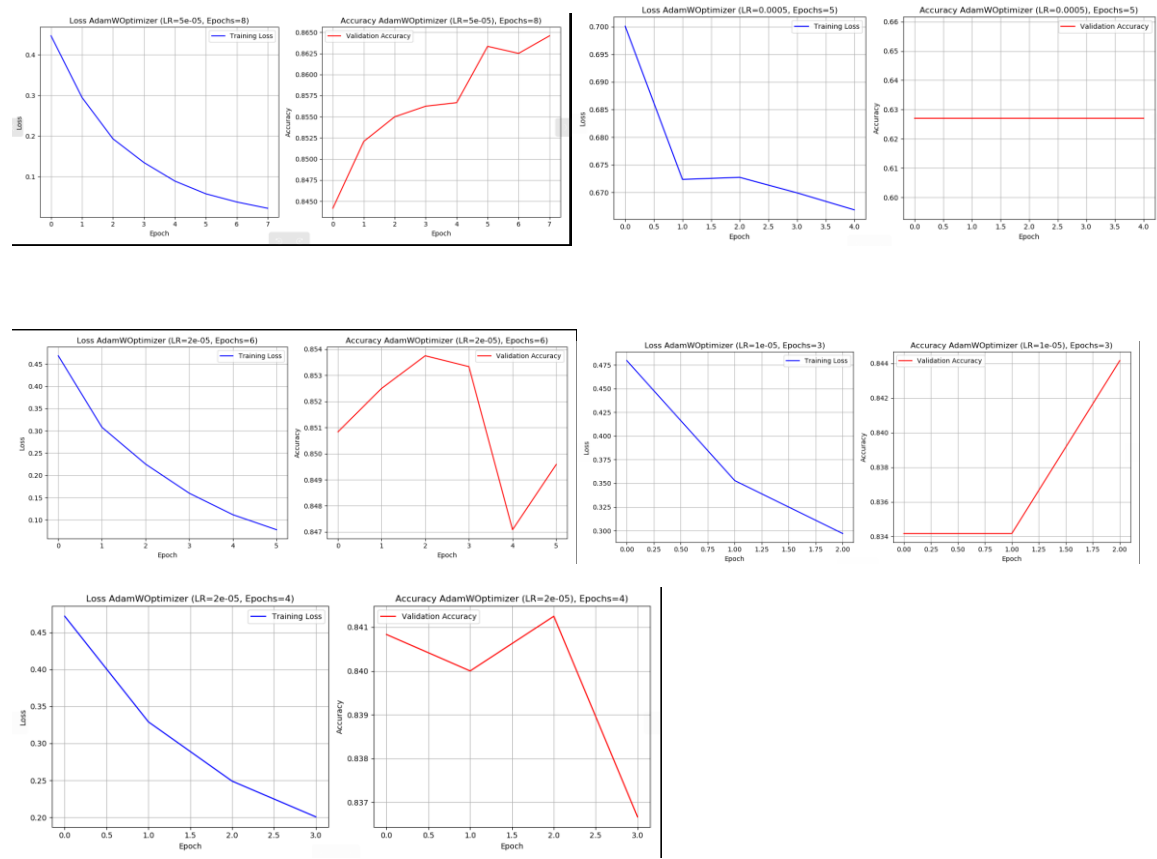
	Bag0	Bag1	Bag2	Bag3	Bag4	Ensemble
Accuracy(f1)	%90,28	%89,98	%90,009	%89,94	%90,002	%90,54

Turkish Movie Sentiment Analysis Dataset

Label	Data Sayısı
4.0	27463
5.0	15873
3.0	14494
2.0	13866
1.0	6381

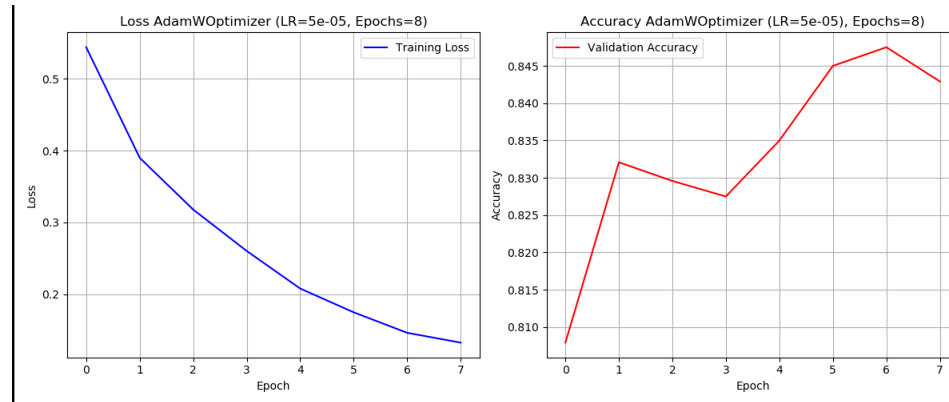
Bu dataset test dosyası içermiyordu o yüzden dosyanın ilk 12000 bilgisiyle bagler oluşturarak train dataseti oluşturdum. Sonrasında da 4000 tane random veriyle bir test dataseti oluşturup tüm modelleri o datasetten geçirdim.

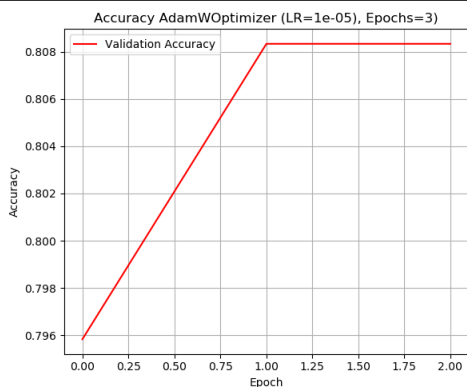
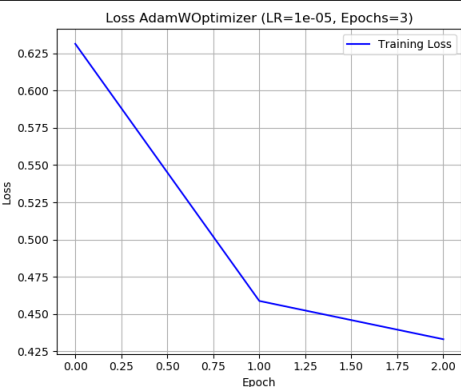
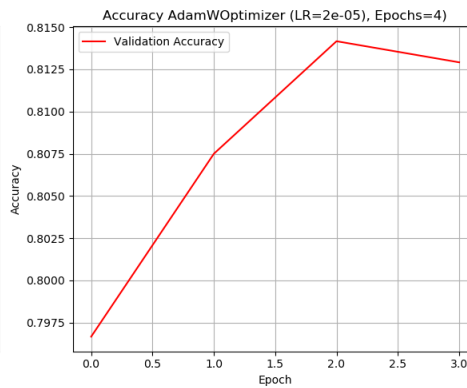
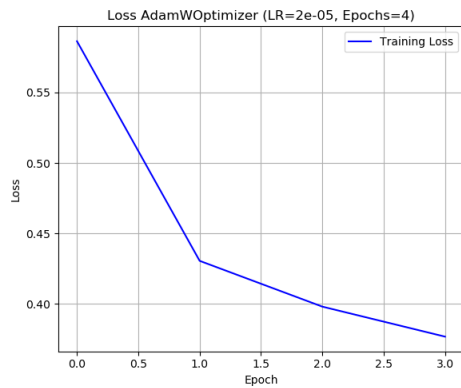
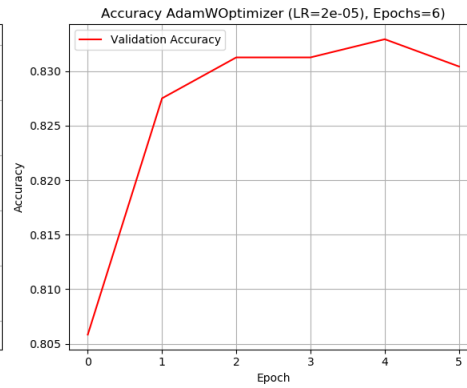
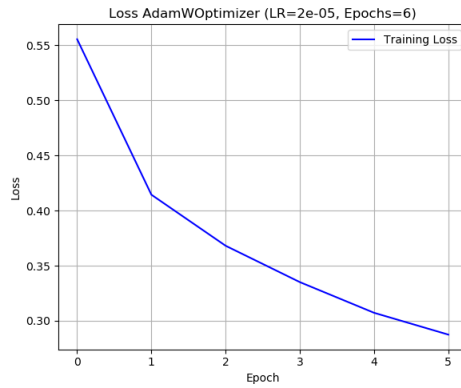
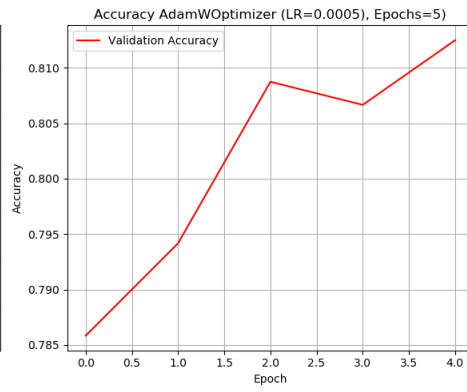
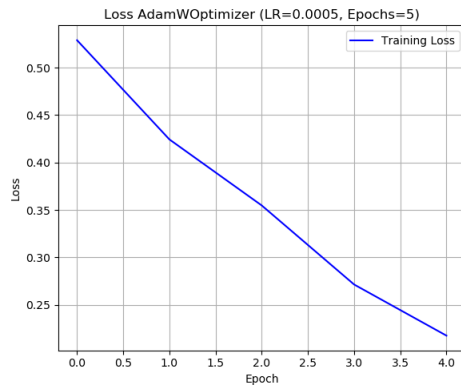
Fine Tuning (Big Bert)



	Bag0	Bag1	Bag2	Bag3	Bag4	Ensemble
Accuracy(f1)	%85,55	%62,75	%84,775	%85,225	%85,5	%86,125

Fine Tuning (Small Bert)

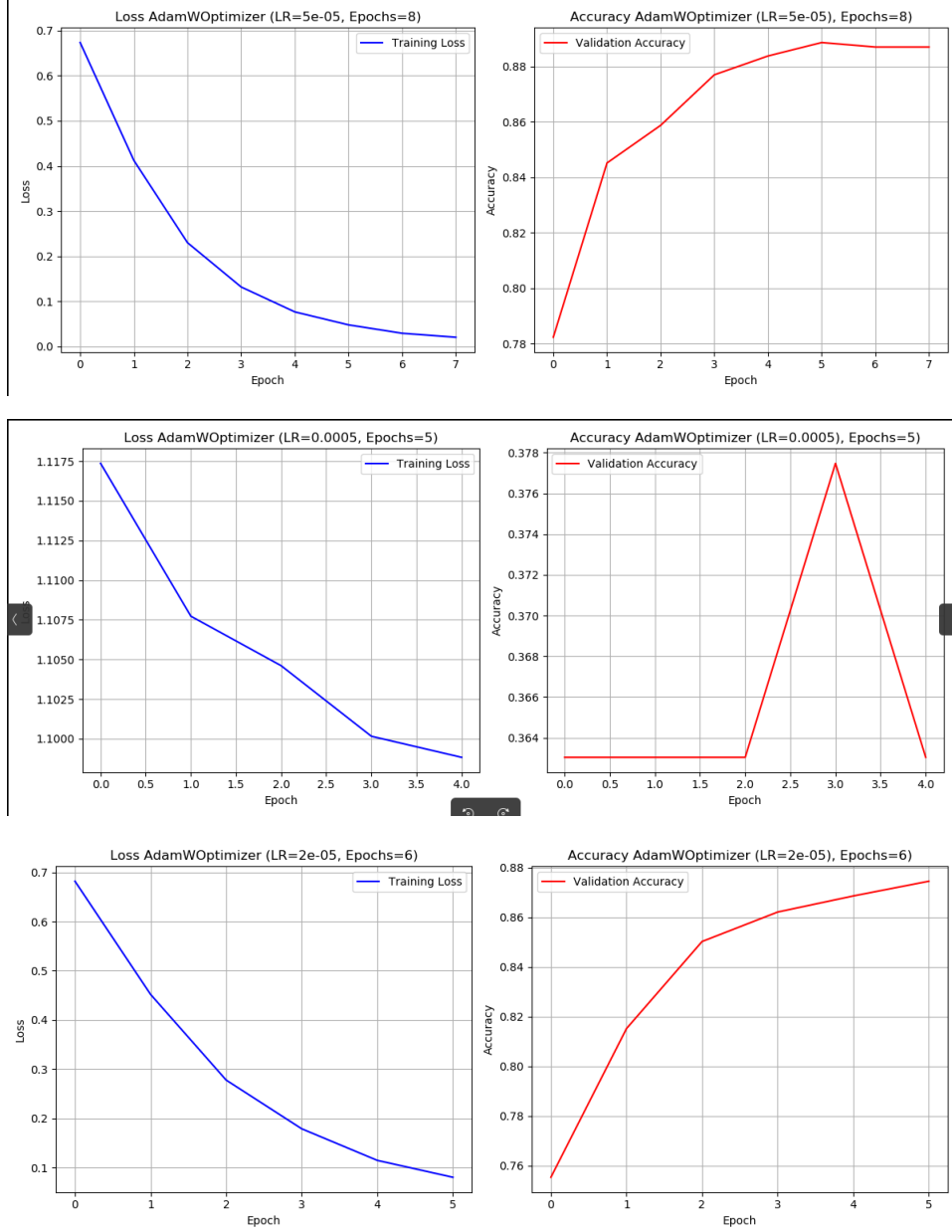


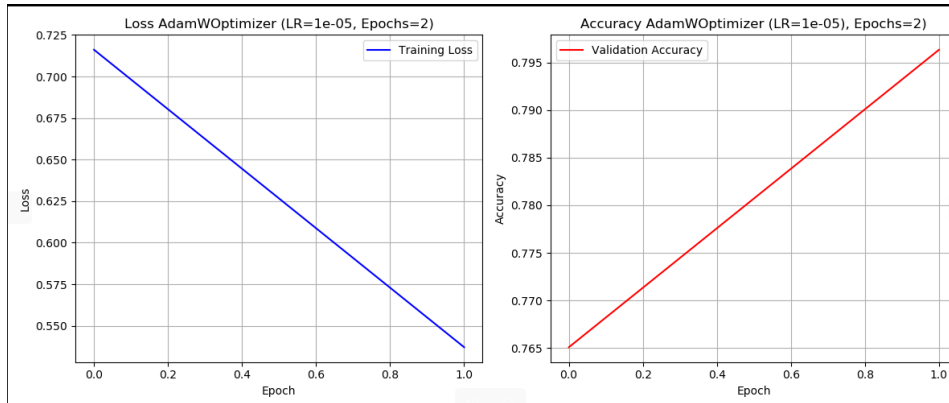
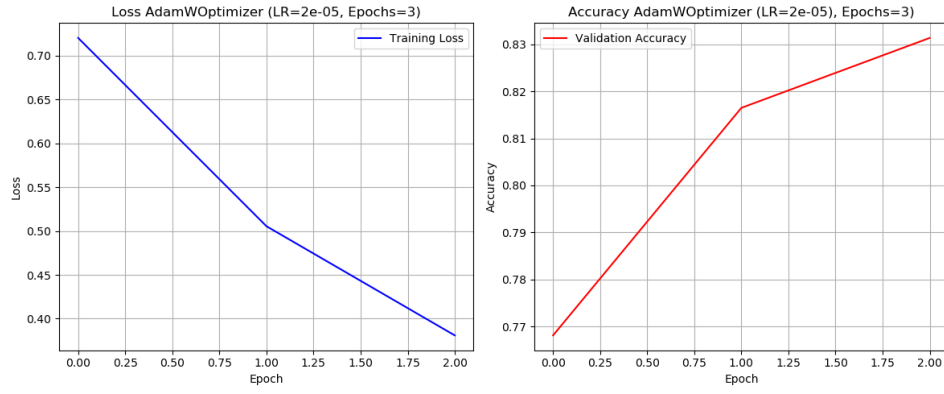


	Bag0	Bag1	Bag2	Bag3	Bag4	Ensemble
Accuracy(f1)	%83,6	%81,125	%82,5	%81,825	%81	%83,1

Duygu Analizi Veri Seti (Olumlu/Olumsuz/Tarafsız)

Label	Data Sayısı
Olumlu	4252
Olumsuz	4252
Tarafsız	2937

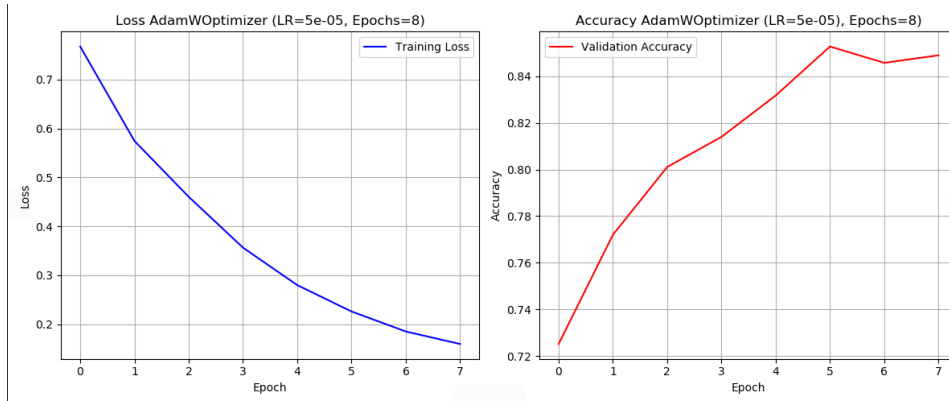
Fine Tuning (Big Bert)



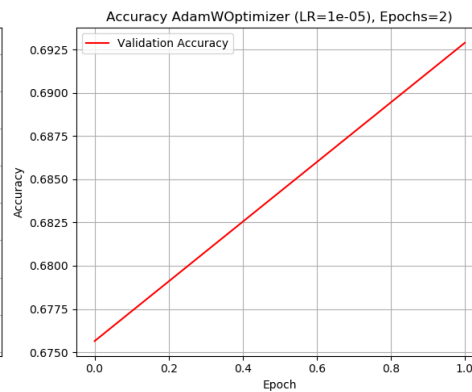
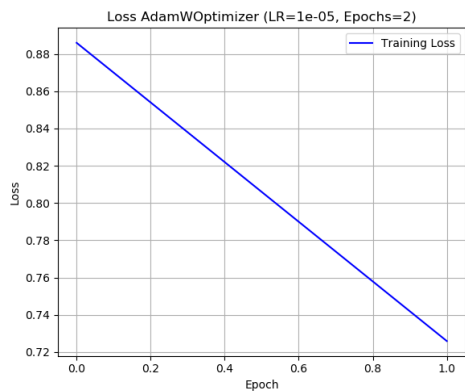
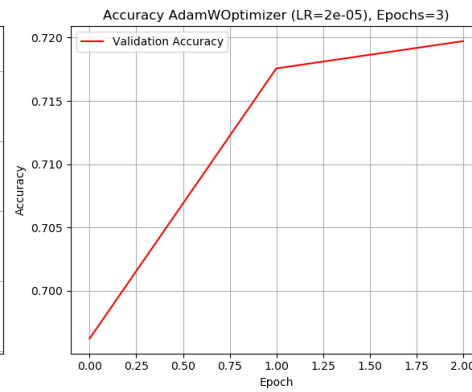
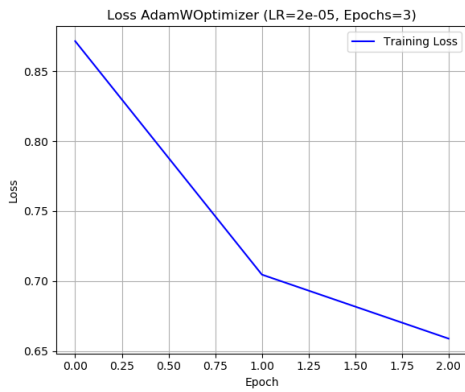
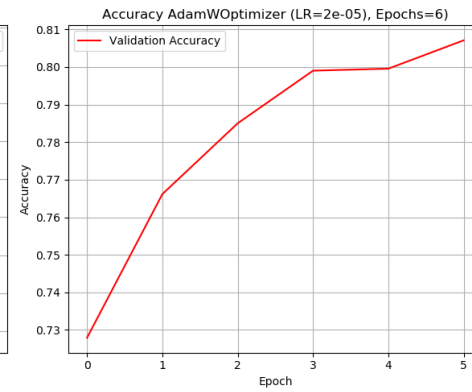
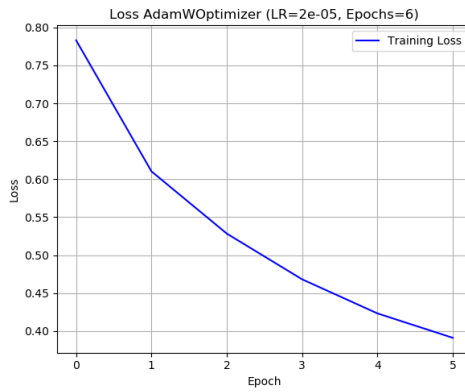
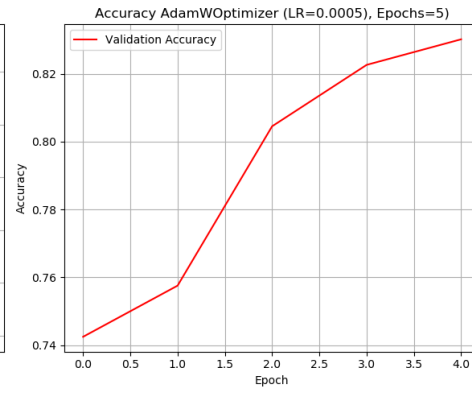
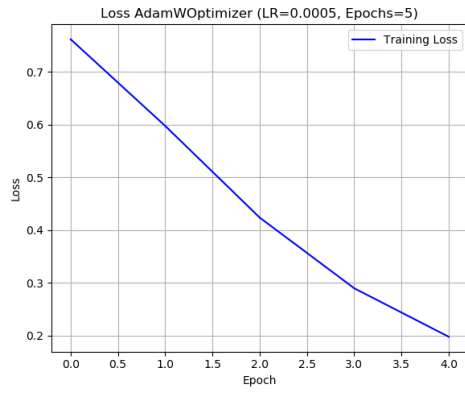
	Bag0	Bag1	Bag2	Bag3	Bag4	Ensemble
Accuracy(f1)	%72,52	%36,37	%73,47	%72,79	%74,15	%74,649
Accuracy(f1)	%72,52	-(Dahil Edilmedi)	%73,47	%73,47	%74,15	%74,78

*Burada Accuracy'lerin birebir gözükmesinin sebebi test sürecinin sonunda oluşan tüm predictionlar ve true labelar pickle dosyasına alınmış başka bir ensemble karar üreticisi dosyasında incelenmiştir.

Fine Tuning (Small Bert)



Alperen Kösem



	Bag0	Bag1	Bag2	Bag3	Bag4	Ensemble
Accuracy(f1)	%71,71	%70,22	%72,39	%71,57	%69,63	%74

Soru B)

- ('RandomForestClassifier1', classifiers[0]),
- ('DecisionTreeClassifier1', classifiers[1]),
- ('SVM', classifiers[2]),
- ('Naive Bayes', classifiers[3]),
- ('K-Nearest Neighbors', classifiers[4]),
- ('GradientBoostingClassifier', classifiers[5]),
- ('DecisionTreeClassifier2', classifiers[6]),
- ('DecisionTreeClassifier3', classifiers[7]),
- ('RandomForestClassifier2', classifiers[8]),
- ('SVM2', classifiers[9]),

Yukarıdaki sınıflandırıcılar kullanılmıştır.

Big Bert Beyaz Perde

RF1	DT1	SVM	KNN	NB	GBC	DT2	DT3	RF2	SVM2
0.7943	0.7831	0.6693	0.6181	0.5275	0.6687	0.785	0.7743	0.7906	0.6643

Accuracy of Ensemble is: 0.89875

Small Bert Beyaz Perde

RF1	DT1	SVM	KNN	NB	GBC	DT2	DT3	RF2	SVM2
0.81	0.7825	0.6525	0.6262	0.5281	0.66	0.785	0.7768	0.7962	0.6418

Accuracy of Ensemble is: 0.8875

Big Bert Movie Sentiment

RF1	DT1	SVM	KNN	NB	GBC	DT2	DT3	RF2	SVM2
0.6719	0.5904	0.6369	0.5294	0.4518	0.6277	0.6014	0.5958	0.6678	0.6375

Accuracy of Ensemble is: 0.66028

Small Bert Movie Sentiment

RF1	DT1	SVM	KNN	NB	GBC	DT2	DT3	RF2	SVM2
0.6677	0.6058	0.6393	0.5380	0.4451	0.6269	0.5956	0.5935	0.6669	0.6379

Accuracy of Ensemble is: 0.6626

Big Bert Magaza Yorum

RF1	DT1	SVM	KNN	NB	GBC	DT2	DT3	RF2	SVM2
0.7548	0.7223	0.5493	0.5347	0.4365	0.5542	0.7098	0.7326	0.7489	0.5368

Accuracy of Ensemble is: 0.7196

Small Bert Magaza Yorum

RF1	DT1	SVM	KNN	NB	GBC	DT2	DT3	RF2	SVM2
0.7668	0.7163	0.5488	0.5189	0.4240	0.5601	0.7201	0.7310	0.7521	0.5298

Accuracy of Ensemble is: 0.7169

Soru C)

Bert Modellerim:

- dbmdz/bert-base-turkish-cased
- ytu-ce-cosmos/turkish-small-bert-uncased
- ytu-ce-cosmos/turkish-medium-bert-uncased
- ytu-ce-cosmos/turkish-tiny-bert-uncased
- ytu-ce-cosmos/turkish-base-bert-uncased

Beyaz Perde

	Model0	Model1	Model2	Model3	Model4	Ensemble
Accuracy(f1)	0.636	0.486	0.516	0.52	0.488	0.632

Yukarıdaki tablodaki labellarım olumlu, olumsuzdu.

Bir de labelları: başta dataset incelemesindeki gibi olumlu cümleler için en çok tekrar eden kelimeyi olumsuz cümleler için en çok tekrar eden kelimeyi alarak yeniden bir çıktı sağladım.

	Model0	Model1	Model2	Model3	Model4	Ensemble
Accuracy(f1)	0.662	0.516	0.568	0.516	0.522	0.564

Magaza Yorumları

	Model0	Model1	Model2	Model3	Model4	Ensemble
Accuracy(f1)	0.388	0.412	0.356	0.346	0.506	0.5

Movie Sentiment

	Model0	Model1	Model2	Model3	Model4	Ensemble
Accuracy(f1)	0.384	0.084	0.044	0.012	0.014	0.04