

Sikayetvar project report

Notes:

1. Accuracy: the fraction of predictions the model got right.
2. Confusion Matrix: a summary table that breaks down the number of correct and incorrect predictions by each class.
3. ROC: a plot that illustrates the true positive rate against the false positive rate at various threshold settings. The area under the curve (AUC) indicates the probability that the classifier will rank a randomly chosen positive observation higher than a randomly chosen negative one.
4. Precision: the fraction of relevant instances among the retrieved instances.
5. Recall: the fraction of the total amount of relevant instances that were actually retrieved.

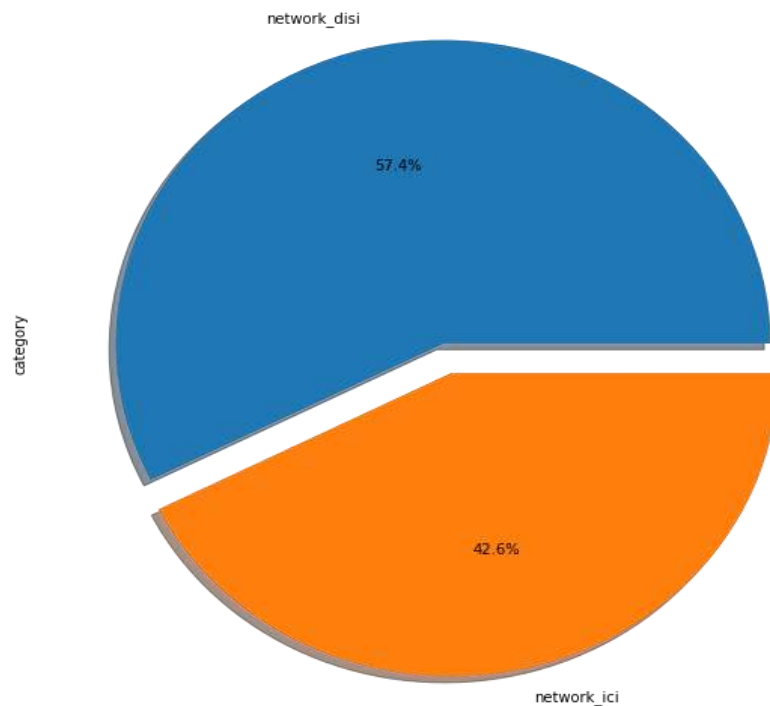
In this report Sekayetva project classification problem into the below steps.

- **This project we used labeled sikayetvar data set, which included 8196 rows sikayetvar data, (4905 rows network out and 3491 rows network in). 0 represent network out, 1 represent network in.**

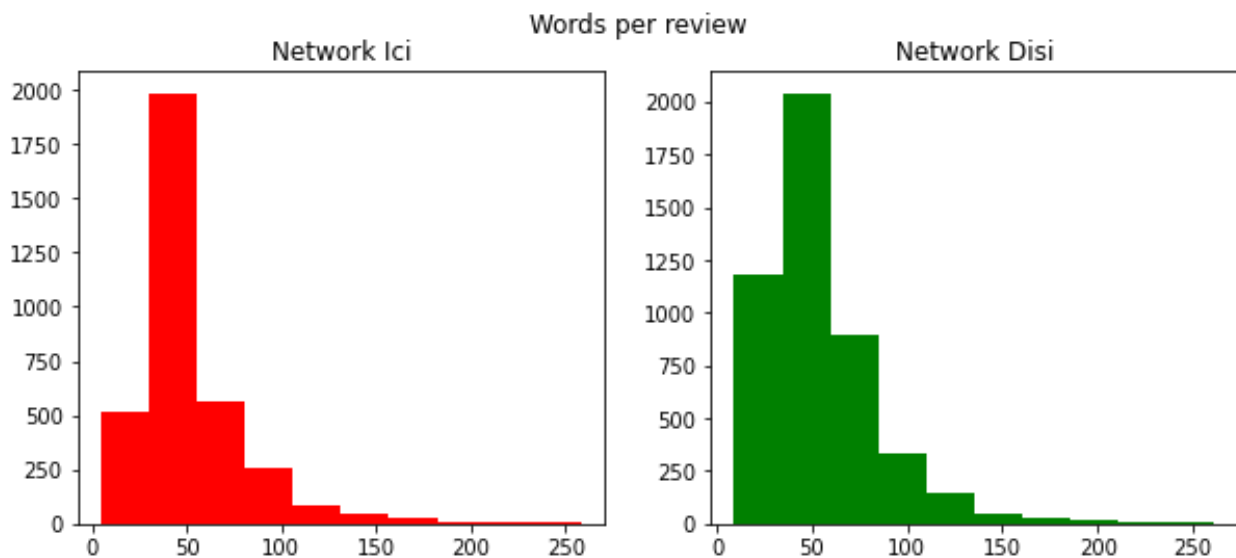
Dataset sample showed below:

	id	baslik	metin	tarih	category
0	17232767	Türk Telekom İnternet Hızım Çok Düşükl	100 lira fatura veriyorum, hala internetim yav...	2021-02-01	1
1	17228655	Türk Telekom Ev İnterneti Sıkıntısı	Aylardır kullanıyoruz Türk Telekom ev internet...	2021-01-31	1
2	17227857	Türk Telekom Port Sorunu Adı Altında İnterneti...	23.11.2020 tarihinde İstanbul Sancaktepe buluş...	2021-01-31	1
3	17226283	Türk Telekom Bağlantı Ve Hız Sorunul	Evde sabit internet sağlayıcı firma Türk Telek...	2021-01-30	1
4	17223192	Türk Telekom'da İnternet Sorunu Yaşıyorum	Türk Telekom sabit internetimde 3 haftadır sor...	2021-01-30	1
5	17214948	Türk Telekom Giderilemeyen İnternet Arızası	Şirketimizin internet hattının hızı 1,5 Mbps y...	2021-01-29	1
6	17209911	Türk Telekom İnternet Çekiyor Hatlar Çekiyor A...	28 Ocak sabah 11'den sonra hatlarda bir sorun ...	2021-01-28	1
7	17212496	Türk Telekom Habersiz Yapılan Taahhüt!	Türk Telekom internet hızımı artırmak için müş...	2021-01-29	0
8	17212086	Türk Telekom Sil Süpürde İnternet Yok	Sil Süpür falan hep boş bana iki aydır haftalı...	2021-01-29	0
9	17239653	Keçpeci Tarafından Koparılan Kablo Ve İlgisiz ...	Yan binamızdaki kentsel dönüşüm var. Bina yıkı...	2021-02-01	1

Dataset sample categories distribution showed below:



Dataset sample plotting with word count per record include network in and network out showed below:



- **Extracting vectors from the “metin” and “baslik”. Most difficult to work with text data while building Machine Learning models. Term Frequency-Inverse Documents Frequencies(TF-IDF) and word Embedding (we use Word2Vec) are two methods we represent in project.**

- Result of running ML algorithms.

Logistic Regression

With tf-idf result

	precision	recall	f1-score	support
0	0.83	0.86	0.84	937
1	0.80	0.76	0.78	703
accuracy			0.82	1640
macro avg	0.82	0.81	0.81	1640
weighted avg	0.82	0.82	0.82	1640

Confusion Matrix: [[806 131]
[168 535]]
AUC: 0.8847218279336462

With word2vec

	precision	recall	f1-score	support
0	0.81	0.81	0.81	953
1	0.74	0.74	0.74	687
accuracy			0.78	1640
macro avg	0.78	0.78	0.78	1640
weighted avg	0.78	0.78	0.78	1640

Confusion Matrix: [[773 180]
[177 510]]
AUC: 0.8465551976368199

Naive Bayes

With tf-idf result

	precision	recall	f1-score	support
0	0.85	0.86	0.85	937
1	0.81	0.79	0.80	703
accuracy			0.83	1640
macro avg	0.83	0.83	0.83	1640
weighted avg	0.83	0.83	0.83	1640

Confusion Matrix: [[805 132]
[147 556]]
AUC: 0.8897513477078719

RandomForestClassifier

With tf-idf result

	precision	recall	f1-score	support
0	0.85	0.86	0.85	937
1	0.81	0.80	0.80	703
accuracy			0.83	1640
macro avg	0.83	0.83	0.83	1640
weighted avg	0.83	0.83	0.83	1640

Confusion Matrix: [[804 133]
[142 561]]
AUC: 0.8949592461640993

With word2vec

	precision	recall	f1-score	support
0	0.78	0.82	0.80	953
1	0.74	0.68	0.71	687
accuracy			0.77	1640
macro avg	0.76	0.75	0.76	1640
weighted avg	0.76	0.77	0.76	1640

Confusion Matrix: [[786 167]
[217 470]]
AUC: 0.8338014788204262

SVM

With tf-idf result

	precision	recall	f1-score	support
0	0.84	0.87	0.86	937
1	0.81	0.79	0.80	703
accuracy			0.83	1640
macro avg	0.83	0.83	0.83	1640
weighted avg	0.83	0.83	0.83	1640

Confusion Matrix: [[811 126]
[149 554]]
AUC: 0.8938320447054929

With word2vec

	precision	recall	f1-score	support
0	0.82	0.82	0.82	953
1	0.75	0.75	0.75	687
accuracy			0.79	1640
macro avg	0.78	0.78	0.78	1640
weighted avg	0.79	0.79	0.79	1640

Confusion Matrix: [[778 175]

[170 517]]

AUC: 0.8523356106740226

- **Output test data**

We select best model in ML running models. In this case I choose SVM with TF-IDF model.

In the result showed if result=1 is network in else network out. Total dataset rows is 221.

	id	baslik	metin	tarih	result
0	19296419	Türk Telekom 1 Yıldır Altyapı İçin Uğraşıyorum	Bir yıldan fazladır binamıza altyapı getirtmey...	2021-10-06	1
1	19296409	Türk Telekom Altyapı Yok Halledilmiyor Sorunu	Bulunmuş olduğum mahallede Esenyurt muhterçesm...	2021-10-06	1
2	19296458	Türk Telekom 566 Tl Haksız Kazanç, Fatura Sorunu	İnternet hizmet numaram;1897967230'infatura hes...	2021-10-06	0
3	19296456	Türk Telekom GPRS Ücreti, Yanlış Bilgi, Fazla ...	Faturam 52 TL ama güncel faturaya baktığımda 1...	2021-10-06	0
4	19296423	Türk Telekom'un Kötü İnternet Altyapısı	İstanbul'un göbeğinde süründüren internet alty...	2021-10-06	1
5	19296405	Türk Telekom Mağduruz	Türk Telekom ev internetini eve bağladım iki t...	2021-10-06	0
6	19296381	Türk Telekom İnternet Hizmetine Ulaşamama	7030337*** numaralı Türk Telekom internet kull...	2021-10-06	1
7	19296312	Türk Telekom Müşteri Hizmetleri Konuşmalarımız...	Türk Telekom cep telefonu taahhüdüm yok ve kon...	2021-10-06	0
8	19296283	Türk Telekom Azdavay Şubesi Sorumsuzluğu	Evimin önüne yapılan inşaat esnasında internet...	2021-10-06	0
9	19296282	Altyapı Sorumsuzluğu Türk Telekom	293/40 sk no.** çamlıpınar Buca/İzmir'de Türk ...	2021-10-06	1
10	19296265	Türk Telekom Evde İnternet Kesintisi	7 gündür altyapı sorunu var denilerek evde tam...	2021-10-06	1
11	19296259	Türk Telekom Modem Teslimatı Sisteme Girilmemiş	23 Haziran 2021' de modemi teslim ettik ancak ...	2021-10-06	0
12	19296217	Türk Telekom Numara Taşıma İşlemi	29.9.2021'de Türk Telekom sistemi üzerinden nu...	2021-10-06	0
13	19296203	Türk Telekom Bağlantı Sorunu Çözüm Yok	02.10.2021'den bu yana internetimiz yok ve Tür...	2021-10-06	1
14	19296163	Türk Telekom Haber Verilmeden Kotalı Pakete Ge...	Türk Telekom ile 1 sene sınırsız internet kont...	2021-10-06	0
15	19296140	Türk Telekom Fatura Fazla Gelecek	Telekom faturalı taahhütlü hattıma 37 lira. Da...	2021-10-06	0
16	19296134	Türk Telekom'un Altyapısı Rezalet Paketim Bile...	Türk Telekom para yüklediğim halde paketimi ta...	2021-10-06	1
17	19296109	Türk Telekom Başansızlıklarını Millete Para Ö...	ADSL 'den VDSL'e geçiş de modemim bozuldu. 3 g...	2021-10-06	0
18	19296035	Türk Telekom İnternet Bağlantımız Bir Haftadır...	Antalya Kepez Fevzi Çakmak mahallesinde Türk T...	2021-10-06	1
19	19296024	Türk Telekom Ev İnternet Bağlantımı Kurmuyor	Günlerdir bir kurum yapılamaz mı! Evi taşıdım ...	2021-10-06	0