



CUKUROVA UNIVERSITY / FACULTY OF ENGINEERING
COMPUTER ENGINEERING

CEN445 – Introduction to Data Visualization

Finding Frequent Patterns Using FP-Growth Algorithm

Asst. Prof. Dr. *****

Tufan Akbaş - *****

Introduction

As technology continues to develop and become more widespread, the amount of data stored in digital environments has grown significantly. However, data in its raw form is often useless. Nevertheless, with appropriate data processing methods, such as data analysis, data mining, and data visualization, it can become a valuable asset in every aspect of life.

Data mining is an analytical method that aims to extract meaningful information from large datasets. The process involves collecting and preparing data, which includes techniques such as data cleaning, organizing missing data, and removing unnecessary information. Then, through data exploration, models suitable for the dataset are found. These models represent patterns in the dataset and can include machine learning algorithms. During the evaluation phase, the models' performance is tested. If the evaluation is successful, the models are put into use.

The FP-Growth algorithm is a method for mining frequent item sets in a dataset. To efficiently generate frequent item sets, the algorithm uses a divide-and-conquer approach. There are several algorithms for finding frequent item sets, such as Apriori, but FP-Growth is a more efficient and faster method. The algorithm scans the database only once and creates an FP-tree.

In our dataset, there are approximately 42.000 news with 13 different categories. All the news are in text(.txt) format. Contents of the news include a title and a paragraph in Turkish.

Purpose and Importance of the Study

Analyzing a dataset of 42,000 news articles using the FP-Growth algorithm is significant for extracting valuable insights and identifying frequent patterns. The primary objective of this study is to uncover meaningful relationships and frequent patterns among the news articles. The FP-Growth algorithm, through association rule analysis, can identify words or topics that frequently co-occur in the news, offering potential benefits such as discovering hidden patterns and improving decision-making processes.

This project can help to identify trends and patterns in news content, making it a valuable resource for shaping marketing strategies and optimizing content production. Furthermore, identifying frequent patterns in customer feedback or areas of interest can contribute to the development of customer-centric strategies for businesses. FP-Growth analysis can reveal hidden information within large datasets, which can strengthen business strategies and enhance decision-making processes with additional insights.

METHOD

On that project, the FP-Growth algorithm is used to find frequent patterns in the 42,000 news dataset. The working principle is realized by creating the FP-Tree, a tree-based structure. In the first step, frequently occurring items in the dataset are identified. Then, the FP-Tree containing these items is created. The FP-Tree represents the frequent patterns in the dataset. When building the FP-Tree, the algorithm compresses the data in such a way that the frequent items are included, and thus efficiently analyzes the data set. Due to this simple structure and efficiency, FP-Growth is used for finding association rules in large data sets.

FP-Growth Algorithm

The FP-Growth algorithm is employed in the analysis of a vast dataset comprising 42,000 news articles distributed across 13 categories. This algorithm proves particularly valuable for revealing frequent patterns within the dataset, assisting in the extraction of meaningful insights. The process begins by establishing a frequency table, where the occurrences of individual words or topics within the news articles are computed, adhering to a predetermined support threshold that ensures a minimum frequency requirement.

Following the creation of the frequency table, the algorithm proceeds to prioritize and sort the words or topics based on their frequency. This meticulous sorting ensures that only those words or topics surpassing the predefined support threshold are considered for further analysis.

The FP-Growth algorithm's core functionality lies in the construction of an FP-Tree tailored to the characteristics of the news articles. This tree structure, formed by encoding each article in a manner that exclusively includes the identified frequent words or topics, serves as a representation of the underlying patterns within the dataset. The iterative construction of the FP-Tree involves sequentially adding each news article to the tree.

As the algorithm advances, it systematically sorts the remaining articles based on the frequency of words or topics and updates the FP-Tree accordingly. This iterative process is applied to the entire dataset, resulting in the creation, and merging of new FP-Trees in each iteration.

The culmination of this process involves merging the generated FP-Trees, yielding a consolidated FP-Tree that encapsulates the collective frequent patterns identified across the diverse categories of news articles. Subsequently, association rules are extracted from this FP-Tree, with additional filtering based on support and confidence values derived from the frequency table.

In essence, the integrated FP-Growth algorithm, operating within the context of 42,000 news articles categorized into 13 distinct categories, offers a powerful and efficient solution for mining frequent patterns and extracting valuable insights from the extensive dataset.

EXPERIMENTAL STUDY

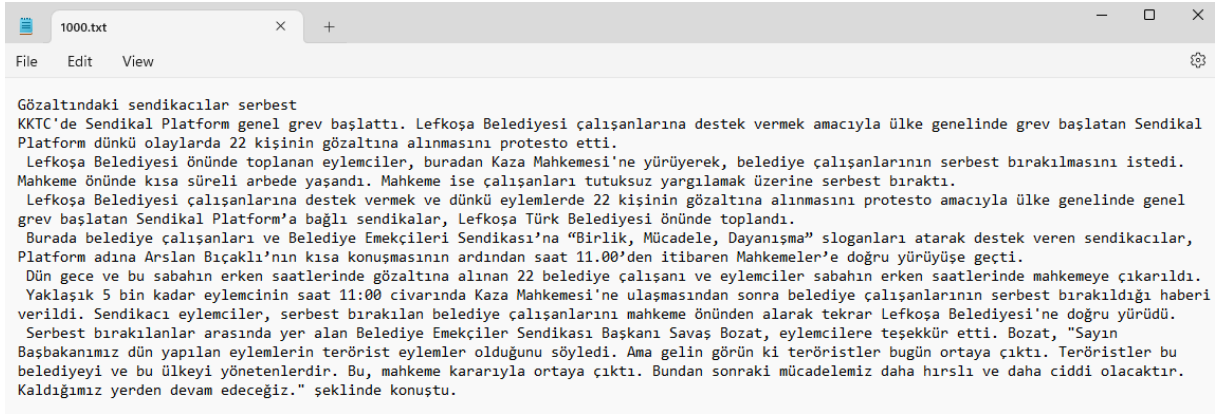


Figure 1

1000.txt,gözetimdeki sendikacılar serbest kktde sendikal platform genel grev başlattı lefkoşa belediyesi çalışanlarına destek vermek amacıyla ülke genelinde grev başlatan sendikal platform çünkü olaylarda kişinin gözetimine alınmasını protesto etti lefkoşa belediyesi önünde toplanan eylemciler buradan kaza mahkemesine yürüyerek belediye çalışanlarının serbest bırakılmasını istedi mahkeme önünde kısa süreli arbeye yaşandı mahkeme çalışanları tutuksuz yargılamak üzerine serbest bıraktı lefkoşa belediyesi çalışanlarına destek vermek çünkü eylemlerde kişinin gözetimine alınmasını protesto amacıyla ülke genelinde genel grev başlatan sendikal platform a bağlı sendikalar lefkoşa türk belediyesi önünde toplandı burada belediye çalışanları belediye emekçileri sendikası na birlik mücadele dayanışma sloganları atarak destek veren sendikacılar platform adına arslan bıçaklı nın kısa konuşmasının ardından saat den itibaren mahkemeler e doğru yürüyüşe geçti dün gece sabahın erken saatlerinde gözetimine alınan belediye çalışanı eylemciler sabahın erken saatlerinde mahkemeye çıkarıldı yaklaşık bin kadar eylemcinin saat civarında kaza mahkemesine ulaşmasından sonra belediye çalışanlarının serbest bırakıldığı haberi verildi sendikacı eylemciler serbest bırakılan belediye çalışanlarını mahkeme önünden alarak tekrar lefkoşa belediyesine doğru yürüdü serbest bırakılanlar arasında yer alan belediye emekçiler sendikası başkanı savaş bozat eylemcilere teşekkür etti bozat sayın başbakanımız dün yapılan eylemlerin terörist eylemler olduğunu söyledi. Ama gelin görün ki teröristler bugün ortaya çıktı. Teröristler bu belediye ve bu ülkeyi yönetenlerdir. Bu, mahkeme kararıyla ortaya çıktı. Bundan sonraki mücadelemiz daha hırslı ve daha ciddi olacaktır. Kaldığımız yerden devam edeceğiz." şeklinde konuştu.

Figure 2

1000.txt,gözetimdeki sendikacılar serbest kktde sendikal platform genel grev başlattı lefkoşa belediyesi çalışanlarına destek vermek amacıyla ülke genelinde grev başlatan sendikal platform çünkü olaylarda kişinin gözetimine alınmasını protesto etti lefkoşa belediyesi önünde toplanan eylemciler buradan kaza mahkemesine yürüyerek belediye çalışanlarının serbest bırakılmasını istedi mahkeme önünde kısa süreli arbeye yaşandı mahkeme çalışanları tutuksuz yargılamak üzerine serbest bıraktı lefkoşa belediyesi çalışanlarına destek vermek çünkü eylemlerde kişinin gözetimine alınmasını protesto amacıyla ülke genelinde genel grev başlatan sendikal platform bağlı sendikalar lefkoşa türk belediyesi önünde toplandı burada belediye çalışanları belediye emekçileri sendikası na birlik mücadele dayanışma sloganları atarak destek veren sendikacılar platform adına arslan bıçaklı nın kısa konuşmasının ardından saat den itibaren mahkemeler doğru yürüyüşe geçti dün gece sabahın erken saatlerinde gözetimine alınan belediye çalışanı eylemciler sabahın erken saatlerinde mahkemeye çıkarıldı yaklaşık bin kadar eylemcinin saat civarında kaza mahkemesine ulaşmasından sonra belediye çalışanlarının serbest bırakıldığı haberi verildi sendikacı eylemciler serbest bırakılan belediye çalışanlarını mahkeme önünden alarak tekrar lefkoşa belediyesine doğru yürüdü serbest bırakılanlar arasında yer alan belediye emekçiler sendikası başkanı savaş bozat eylemcilere teşekkür etti bozat sayın başbakanımız dün yapılan eylemlerin terörist eylemler olduğunu söyledi gelin görün teröristler bugün ortaya çıktı teröristler belediye ülkeyi yönetenlerdir mahkeme kararıyla ortaya çıktı bundan sonraki mücadelemiz hırslı ciddi olacaktır kaldığımız yerden devam edeceğiz şeklinde konuştu

Figure 3

1000.txt,gözetimdeki sendikacılar serbest kktde sendikal platform genel grev başlattı lefkoşa belediyesi çalışanlarına destek vermek amacıyla ülke genelinde grev başlatan sendikal platform çünkü olaylarda kişinin gözetimine alınmasını protesto etti lefkoşa belediyesi önünde toplanan eylemciler buradan kaza mahkemesine yürüyerek belediye çalışanlarının serbest bırakılmasını istedi mahkeme önünde kısa süreli arbeye yaşandı mahkeme çalışanları tutuksuz yargılamak üzerine serbest bıraktı lefkoşa belediyesi çalışanlarına destek vermek çünkü eylemlerde kişinin gözetimine alınmasını protesto amacıyla ülke genelinde genel grev başlatan sendikal platform bağlı sendikalar lefkoşa türk belediyesi önünde toplandı burada belediye çalışanları belediye emekçileri sendikası birlik mücadele dayanışma sloganları atarak destek veren sendikacılar platform adına arslan bıçaklı nın kısa konuşmasının ardından saat den itibaren mahkemeler doğru yürüyüşe geçti dün gece sabahın erken saatlerinde gözetimine alınan belediye çalışanı eylemciler sabahın erken saatlerinde mahkemeye çıkarıldı yaklaşık bin kadar eylemcinin saat civarında kaza mahkemesine ulaşmasından sonra belediye çalışanlarının serbest bırakıldığı haberi verildi sendikacı eylemciler serbest bırakılan belediye çalışanlarını mahkeme önünden alarak tekrar lefkoşa belediyesine doğru yürüdü serbest bırakılanlar arasında yer alan belediye emekçiler sendikası başkanı savaş bozat eylemcilere teşekkür etti bozat sayın başbakanımız dün yapılan eylemlerin terörist eylemler olduğunu söyledi gelin görün teröristler bugün ortaya çıktı teröristler belediye ülkeyi yönetenlerdir mahkeme kararıyla ortaya çıktı bundan sonraki mücadelemiz hırslı ciddi olacaktır kaldığımız yerden devam edeceğiz şeklinde konuştu

Figure 4

1000.txt,göztaltındaki sendikacılar serbest kktde sendikal platform genel grev bařlattı lefkořa belediyesi alıřanlarına destek vermek amacıyla lke genelinde grev bařlatan sendikal platform dnk olaylarda kiřinin gztaltına alınmasını protesto etti lefkořa belediyesi nnde toplanan eylemciler buradan kaza mahkemesine yryerek belediye alıřanlarının serbest bırakılmasını istedi mahkeme nnde kısa sreli arbeye yařandı mahkeme alıřanları tutuksuz yargılamak zerine serbest bıraktı lefkořa belediyesi alıřanlarına destek vermek dnk eylemlerde kiřinin gztaltına alınmasını protesto amacıyla lke genelinde genel grev bařlatan sendikal platform baęlı sendikalar lefkořa trk belediyesi nnde toplandı burada belediye alıřanları belediye emekileri sendikası birlik mcadele dayanıřma sloganları atarak destek veren sendikacılar platform adına arslan bıaklı nın kısa konuřmasının ardından saat den itibaren mahkemeler doęru yryře geti dn gece sabahın erken saatlerinde gztaltına alınan belediye alıřanı eylemciler sabahın erken saatlerinde mahkemeye ıkarıldı yaklaşık bin eylemcinin saat civarında kaza mahkemesine ulařmasından sonra belediye alıřanlarının serbest bırakıldığı haberi verildi sendikacı eylemciler serbest bırakılan belediye alıřanlarını mahkeme nnden alarak tekrar lefkořa belediyesine doęru yrd serbest bırakılanlar arasında yer alan belediye emekiler sendikası bařkanı savař bozat eylemcilere teřekkr etti bozat sayın bařbakanımız dn yapılan eylemlerin terrist eylemler olduęunu syledi gelin grn terristler bugn ortaya ıktı terristler belediyeyi lkeyi ynetenlerdir mahkeme kararıyla ortaya ıktı bundan sonraki mcadelemiz hırslı ciddi olacaktır kaldığımız yerden devam edeceęiz řeklinde konuřtu

Figure 5

1000.txt,"gztaltındaki,sendikacılar,serbest,kktde,sendikal,platform,genel,grev,bařlattı,lefkořa,belediyesi,alıřanlarına,destek,vermek,amacıyla,lke, genelinde,grev,bařlatan,sendikal,platform,dnk,olaylarda,kiřinin,gztaltına,alınmasını,protesto,etti,lefkořa,belediyesi,nnde,toplanan,eylemciler, buradan,kaza,mahkemesine,yryerek,belediye,alıřanlarının,serbest,bırakılmasını,istedi,mahkeme,nnde,kısa,sreli,arbeye,yařandı,mahkeme,alıřanları, tutuksuz,yargılamak,zerine,serbest,bıraktı,lefkořa,belediyesi,alıřanlarına,destek,vermek,dnk,eylemlerde,kiřinin,gztaltına,alınmasını,protesto, amacıyla,lke,genelinde,genel,grev,bařlatan,sendikal,platform,baęlı,sendikalar,lefkořa,trk,belediyesi,nnde,toplandı,burada,belediye,alıřanları, belediye,emekileri,sendikası,birlik,mcadele,dayanıřma,sloganları,atarak,destek,veren,sendikacılar,platform,adına,arslan,bıaklı,nın,kısa,konuřmasının, ardından,saat,den,iterbaren,mahkemeler,doęru,yryře,geti,dn,gece,sabahın,erken,saatlerinde,gztaltına,alınan,belediye,alıřanı,eylemciler,sabahın, erken,saatlerinde,mahkemeye,ıkarıldı,yaklaşık,bin,eylemcinin,saat,civarında,kaza,mahkemesine,ulařmasından,sonra,belediye,alıřanlarının,serbest, bırakıldığı,haberi,verildi,sendikacı,eylemciler,serbest,bırakılan,belediye,alıřanlarını,mahkeme,nnden,alarak,tekrar,lefkořa,belediyesine,doęru,yrd, serbest,bırakılanlar,arasında,yer,alan,belediye,emekiler,sendikası,bařkanı,savař,bozat,eylemcilere,teřekkr,etti,bozat,sayın,bařbakanımız,dn,yapılan, eylemlerin,terrist,eylemler,olduęunu,syledi,gelin,grn,terristler,bugn,ortaya,ıktı,terristler,belediyeyi,lkeyi,ynetenlerdir,mahkeme,kararıyla, ortaya,ıktı,bundan,sonraki,mcadelemiz,hırslı,ciddi,olacaktır,kaldığımız,yerden,devam,edeceęiz,řeklinde,konuřtu"

Figure 6

0.2313810556760665, olarak, eden
0.2284887924801157, olduğunu, eden
0.2031814895155459, eden, olan
0.2031814895155459, söyledi, eden
0.21475054229934923, ifade, eden
0.25090383224873464, olduğunu, ifade
0.22270426608821403, olarak, ifade
0.21691973969631237, söyledi, ifade
0.24945770065075923, olarak, fazla
0.23355025307302965, fazla, olan
0.22631959508315258, olduğunu, fazla
0.22270426608821403, olarak, sağlık
0.24150397686189443, olarak, tedavi
0.23427331887201736, tedavi, olan
0.30874909616775126, önemli, olarak
0.28054953000723065, önemli, olan
0.20824295010845986, önemli, olduğu
0.27042660882140274, olduğunu, önemli
0.2198120028922632, önemli, olarak, olan
0.20607375271149675, olduğunu, önemli, olarak
0.2299349240780911, olarak, özellikle
0.2284887924801157, özellikle, olan
0.21258134490238612, ortaya, olarak
0.21402747650036152, son, olarak

1	Word, Frequency
2	ortak, 2259
3	vizyonumuz, 16
4	var, 17209
5	dışışleri, 2404
6	bakanı, 5944
7	davutoğlu, 861
8	yunanistan, 643
9	türkiye, 16225
10	arasındaki, 2553
11	farlılıkların, 1
12	vizyon, 99
13	çözülebileceğini, 18
14	söyledi, 14956
15	ahmet, 3581
16	yunan, 564
17	mevkidaşı, 54
18	dimitris, 72
19	avramopulos, 35
20	başbaşa, 40
21	heyetler, 46
22	arası, 1079
23	görüşmelerin, 413
24	ardından, 12890
25	basın, 3901
26	toplantısı, 890
27	düzenledi, 505
28	yunanistanda, 122
29	kendisine, 2519
30	gösterilen, 507
31	sıcak, 954
32	karşılama, 117
33	teşekkür, 1728
34	ederek, 3180
35	ziyaretinin, 117
36	türkyunan, 47
37	dostluğunun, 12
38	nişanesi, 4
39	olduğunu, 19728

1	support,itemsets
2	0.203182,yer
3	0.208206,tarafından
4	0.208349,devam
5	0.208921,oldu
6	0.209659,yapılan
7	0.210778,yeni
8	0.212683,eden
9	0.21485,önce
10	0.221852,ardından
11	0.228401,ilgili
12	0.23926,olduğu
13	0.243689,büyük
14	0.245332,söyledi
15	0.252405,etti
16	0.256644,son
17	0.257311,ilk
18	0.267741,olduğunu
19	0.294127,sonra
20	0.306225,dedi
21	0.363641,olan
22	0.400957,olarak
