

ÖZET

Veri madenciliği bir veri ambarında, veriler üzerindeki gizli bağlantıları ortaya çıkarmak için kullanılmaktadır. Verinin bulunduğu her ortamda ham veriler anlamlı hale getirilerek bilgi keşfi işlemine veri madenciliği denir.

Günümüzde ilerleyen teknoloji ve bilgiye erişiminin artması ile birlikte kurum ve kuruluşlar bu bilgilerden faydalanarak üretim, satış ve kaynakların yönetilmesi gibi konularda daha doğru, düzgün ve yeni kararlar verebilmek için önemli bir yol olarak veri madenciliğinden yararlanmanın faydalı olabileceğini düşünmüşlerdir. Bununla birlikte veri madenciliğinde yeni yöntemler, algoritmalar, düşünceler gelişmiş ve sektörlerin için önem arz eden bir durum haline gelmiştir.

Bilgi miktarının büyük oranlarda arttığı bu bilgi çağında büyük hacimlerdeki verilerden anlamlı bilgilerin elde edilmesi bir süreç gerektirmektedir. Bu sürecin en önemli adımı ise veri madenciliğidir.

Bu çalışmada bilginin ortaya çıkarılması süreci araştırılmış ve incelenmiştir. Yine aynı şekilde sürecin en önemli adımı olan veri madenciliği adımı da ayrıntılı olarak incelenmiştir.

Tez içerisinde anlamlı bilgiye ulaşma süreci aşamalarıyla açıklanmıştır. Sürecin en önemli adımı olan veri madenciliği adımı ayrıntılı olarak ele alınmış ve veri madenciliği tekniklerinden bahsedilmiştir. Veri madenciliği tekniklerinden olan kümeleme analizi ayrıntılı olarak incelenmiştir.

Tezin son kısmında veri madenciliği algoritmalarından biri olan k-means algoritması sigorta veri tabanına uygulanmış ve sonuçlar ortaya konulmuştur.

ABSTRACT

Data mining is used to find hidden connections on the data via data warehouse. Knowledge discovery means gathering meaningful data, on any environment where the data is found, the process is called data mining.

With the increasing technology and access to information nowadays, institutions and organizations have benefited from this information and thought to benefit from data mining as an important way to make more accurate, smooth and new decisions on issues such as production, sales and management of resources. In addition, new methods, algorithms, ideas developed in data mining have become important for the sectors.

In this information age, where the amount of information has increased greatly, it is necessary to obtain meaningful information from large volumes of data. The most important step of this process is data mining.

In this study, the process of revealing information has been researched and examined. Likewise, the most important step of the process, the data mining step, has also been examined in detail.

The process of reaching meaningful information in the thesis is explained with the stages. Data mining step, which is the most important step of the process, has been discussed in detail and data mining techniques are mentioned. Clustering analysis, which is one of the data mining techniques, has been examined in detail.

In the last part of the thesis, k-means algorithm, which is one of the data mining algorithms, is applied to the insurance database and the results are presented.