



Düzce University Journal of Science & Technology

Research Article

Performance Comparison of Association Rule Algorithms with SPMF on Automotive Industry Data

 Melih NAIR^{a,*},  Fatih KAYAALP^b

^a Department of Computer Engineering, Institute of Natural Sciences, Düzce University, Düzce, TURKEY

^b Department of Computer Engineering, Faculty of Engineering, Düzce University, Düzce, TURKEY

* Corresponding author's e-mail address: melihnair@gmail.com

DOI : 10.29130/dubited.581931

ABSTRACT

By the recent developments about the information technologies, companies can store their data faster and easier with lower costs. All transactions (sales, current card, invoicing, etc.) performed in companies during the day combine at the end of the day to form big datasets. It is possible to extract valuable information through these datasets with data mining. And this has become more important for companies in terms of today's conditions where the competition in the market is high. In this study, a dataset of a company selling car maintenance and repair products in Turkey is used. Association Rules are applied on this dataset for determining the items which are bought together by the customers. These rules, which are calculated specifically for the company, can be used to redefine the sales and marketing strategies, to revise the storage areas efficiently, and to create sales campaigns suitable for the customers and regions. These algorithms are also called Frequent Itemset Mining Algorithms. The most recent 11 algorithms from these are applied to this dataset in order to compare the performances according to metrics like memory usage and execution times against varying support values and varying record numbers by using SPMF platform. Three different datasets are created by using the whole dataset like 6-months, 12-months and 22-months. According to the experiments, it can be said that execution times generally increases inversely with the support values as nearly all algorithms have higher execution time values for the lowest support value of 0.1. dEclat_bitset algorithm has the most efficient performance for 6-months and 12-months dataset. But Eclat algorithm can be said to be the most efficient algorithm for 0.7 and 0.3 support values; on the other hand dEclat_bitset is the most efficient algorithm for 0.3 and 0.1 support values on 22-months dataset.

Keywords: Association Rules, Market Basket Analysis, SPMF.

Otomotiv Endüstrisi Verileri Üzerinde Birliktelik Kuralları Algoritmalarının SPMF ile Performans Karşılaştırması

ÖZET

Bilgi teknolojilerindeki son gelişmeler sayesinde, şirketler verilerini daha düşük maliyetlerle daha hızlı ve daha kolay saklayabilirler. Gün içinde şirketlerde gerçekleştirilen tüm işlemler (satışlar, cari kartlar, faturalama vb.), günün sonunda birleştirilir ve büyük veri setleri oluştururlar. Bu veri setlerinden veri madenciliği aracılığıyla değerli

bilgiler elde edilmesi mümkündür. Pazardaki rekabetin yüksek olduğu günümüz şartları açısından bu durum şirketler için çok daha önemli hale gelmiştir. Bu çalışmada Türkiye’de araç bakım ve servis ürünleri satan bir şirketin veriseti kullanılmıştır. Bu verisetine, müşteriler tarafından birlikte satın alınmış olan ürünlerin tespiti için Birliktelik Kuralları uygulanmıştır. Şirketlere özgü olarak çıkarımı yapılan bu kurallar şirketlerin satış ve pazarlama stratejilerinin belirlenmesinde, depoların verimli bir şekilde kullanımlarında ve müşteriler ya da bölgelere göre uygun satış kampanyaları oluşturulmasında kullanılabilir. Birliktelik kuralları aynı zamanda Sık Satılan Ürün Algoritmaları olarak da isimlendirilebilmektedir. Bu algoritmalarından en güncel 11 tanesi SPMF yazılımı kullanılarak bu veri setine uygulanmış ve bu algoritmaların değişken destek değerleri ve değişken kayıt sayılarına bağlı olarak performansları, bellek kullanım miktarları ve işlem süreleri açısından karşılaştırılmıştır. Başlangıçtaki veri seti, 6 aylık, 12 aylık ve 22 aylık kayıt içerecek şekilde 3 ayrı veri seti haline getirilmiştir. Deney sonuçlarına bakıldığında, işlem zamanlarının genellikle destek değerleriyle ters orantılı olarak arttığı söylenebilir. Çünkü neredeyse tüm algoritmaların en düşük destek değeri olan 0,1 için daha yüksek işlem zamanı değerlerine sahip oldukları görülmüştür. 6 aylık ve 12 aylık veri setleri için dEclat_bitset algoritması en verimli performansı göstermiştir. Fakat 22 aylık veri setinde, 0,7 ve 0,3 destek değerleri için Eclat algoritması en verimli olarak görünürken; 0,3 ve 0,1 destek değerleri için dEclat_bitset algoritması en verimli olarak görünmektedir.

Anahtar Kelimeler: Birliktelik Kuralları, Market Sepeti Analizi, SPMF.