

BAYESCI AĞLARDA RİSK ANALİZİ: BANKACILIK SEKTÖRÜ ÜZERİNE BİR UYGULAMA

RISK ANALYSIS IN BAYESIAN NETWORKS: AN APPLICATION ON BANKING SECTOR

*Emre DÜNDER**

*Mehmet Ali CENGİZ***

*Haydar KOÇ****

*Nurettin SAVAŞ*****

ÖZET

Bayesci ağlar çok değişkenli bir veri seti içindeki belirsiz ilişkisel yapıları grafiksel modeller ile betimleyen matematiksel nesnelere dir. Bayesci ağlar çıkarsama açısından elverişli olduğu için birçok farklı alanda yaygın biçimde kullanılmaktadır. Özellikle son yıllarda sigortacılık, risk yönetimi ve suistimal tespiti konularında Bayesci ağlar uygulanmaktadır. Bir Bayesci ağı oluşturmak için yapı öğrenme ve parametre öğrenme algoritmaları kullanılmaktadır. Bu çalışmada risk analizi üzerine bir veri seti kullanılarak oluşturulan Bayesci ağ yapılarının performansı karşılaştırılmıştır. Bayesci ağ yapısı oluşturulduktan sonra koşullu olasılık değerleri hesaplanmış ve bu değerlere göre kredi risk durumu ve diğer değişkenlere ilişkin değerlendirmeler yapılmıştır. Uygulamalar GeNIe programı ve R programı içerisinde bulunan bnlearn paketi kullanılarak gerçekleştirilmiştir.

Anahtar Kelimeler: Risk analizi, Bayesci ağlar, yapı öğrenme, parametre öğrenme

ABSTRACT

Bayesian Networks are the mathematical objects that represent uncertain relational structures in a multivariate dataset. Bayesian networks are widely used in many fields for being suitable in terms of inference. Especially in recent years Bayesian Networks are implemented such as in insurance, risk management and fraud detec-

* Arş. Gör. Ondokuz Mayıs Üniv., Fen Ed. Fak. İstatistik Böl., emre.dunder@omu.edu.tr

** Doç. Dr. Ondokuz Mayıs Üniv., Fen Ed. Fak. İstatistik Böl., macengiz@omu.edu.tr

*** Arş. Gör. . Ondokuz Mayıs Üniv., Fen Ed. Fak. İstatistik Böl., haydarkoc@omu.edu.tr

**** Yrd. Doç. Dr. Erzincan Üniversitesi, İ.İ.B.F. İşletme Böl. nuretins24@hotmail.com

tion. To construct a Bayesian network, structure and parameter learning algorithms are employed. In this study the performance of constructed Bayesian networks structures are compared using a dataset on risk analysis. After constructing the Bayesian network structure the conditional probability values are calculated and according to these values, the assessments are made about credit risk position and the other variables. The applications are performed using GenIe and bnlearn package existing in R programme.

Keywords: Risk analysis, Bayesian networks, structure learning, parameter learning

1.GİRİŞ

Bayesci ağlar belirsiz ilişkileri ortaya çıkarmak üzere matematiksel bir model sunar. (Gemela, 2001). Bayesci ağlar değişkenler arası ilişkileri belirlemek için grafiksel modellerden yararlanmaktadır. Bayesci ağlar günümüzde tıp, biyoloji, psikoloji gibi birçok farklı alanda kullanılmaktadır. Özellikle risk analizi alanında risk değerlerini tahmin etmek üzere Bayesci ağlar etkin bir biçimde kullanılmaktadır (Oliva ve ark., 2009; Fenton ve Neil, 2012). Bankacılık alanında kredi risk durumuna ilişkin bir veri seti içerisinde bulunan değişkenler kullanılarak, risk düzeyini etkileyen nedensel faktörlerin belirlenmesi ve risk değerlerinin hesaplanmasına yönelik Bayesci ağ yapıları oluşturulmuş ve alternatif bir model oluşturulacaktır.

2.BAYESCİ AĞLAR

Bayesci ağlar, çok değişkenli bir veri kümesi için değişkenler arası nedensellik ve koşullu bağımsızlık ilişkilerini grafiksel olarak betimleyen grafiksel modellerdir. Bir Bayesci ağı oluşturan üç temel bileşen vardır:

1. $X = \{ X_1, X_2, \dots, X_n \}$ şeklindeki rastlantı değişkenleri kümesi
2. $G = (V, E)$ biçiminde ifade edilen yönlü döngüsel olmayan bir grafik (Y.D.G) .
3. P ile gösterilen, yerel olasılık dağılımlarının çarpımsal biçimiyle ifade edilen bir ortak olasılık dağılımı.

Bayesci ağlarda olasılık teorisinden ortak olasılık dağılımını oluşturma ve koşullu olasılıkların elde edilmesinde yararlanır. Ebeveynleri belirlenmiş olan değişkenler için koşullu olasılık tabloları oluşturulur. Bayesci ağlarda değişkenlerin yerel olasılık dağılımlarından yola çıkarak bir ortak olasılık dağılımı kurulur. Bağımlılık yapılarının görsel olarak gösteriminde grafiksel modeller kullanılır. Grafiksel modeller aracılığı ile değişkenler arası ilişkiler ve koşullu bağımsızlık durumları hakkında çıkarımlar yapılabilir. Bayesci ağ yapısının meydana getirilmesi sürecinde çeşitli paket programlar kullanılır. Özellikle öğrenme algoritmalarının yazılımında programlama dillerinden

etkin bir şekilde yararlanılmaktadır. İstatistik teorisi Bayesci ağların oluşturulmasında çok önemli bir konuma sahiptir. Bayesci ağ yapısının içinde bulunan mevcut değişkenler arası ilişkilerin belirlenmesi sürecinde çeşitli istatistiksel testler uygulanır. Ayrıca parametrelerin öğrenilmesi aşamasında koşullu olasılık ve Bayesci yöntemlerden faydalanılır.

2.1. Bayesci Ağlarda Öğrenme Algoritmaları

Yapı öğrenme algoritmaları, Bayesci ağın grafiksel yapısını kendiliğinden oluşturmak üzere kullanılan yöntemlerdir. Yapı öğrenme algoritmaları konusunda ilk çalışma (Verma ve Pearl, 1991) tümevarımsal nedensellik algoritması üzerinde durulmuştur. Bu algoritma koşullu bağımsızlık testlerini kullanarak Bayesci ağların yapısını öğrenmek için bir altyapı sağlamaktadır (Scutari, 2011). Kısıtlama tabanlı algoritmalar için Spirtes ve ark. (2000), tümevarımsal nedensellik algoritmasını pratik bir şekilde uygulayabilmeyi sağlayan PC algoritmasını geliştirmişlerdir. Markov örtülerinin elde edilmesine dayalı olarak çeşitli kısıtlama tabanlı algoritmalar geliştirilmiştir. Margaritis (2003), değişkenlerin ilk olarak Markov örtülerine dayalı Büyüme-Daralma isimli bir algoritma önermiştir. Tsamardinos ve ark. (2003) iki aşamadan oluşan Artımsal İlişki Markov Örtüsü ve Aralıklı Artımsal İlişki Markov Örtüsü algoritmalarını geliştirmiştir. Yaramakala ve ark. (2005) koşullu bağımsızlı ilişkilerini G^2 test istatistiğini kullanarak test eden Hızlı Artımsal İlişki Markov Örtüsü algoritmasını tanıtmıştır. Tsamardinos ve ark. (2003) aday ebeveyn ve çocuklar adını verdikleri bir kümeden yararlanarak, iki aşamadan oluşan En Çok-En Az-Ebeveynler ve Çocuklar algoritmasını tanıtmıştır.

Bir Bayesci ağı oluştururken yapı ve parametre olmak üzere iki temel kısım elde edilmelidir. Bayesci ağın yönlü döngüsüz grafiği yapı, koşullu olasılık değerleri parametre olarak adlandırılır (Neapolitan, 2003). Bir Bayesci ağı oluşturma sürecine öğrenme denir. Bayesci ağlarda öğrenme yapı ve parametre öğrenme olmak üzere ikiye ayrılır. Yapı öğrenme, çok değişkenli bir veri setinden yararlanılarak değişkenler arası ilişkileri gösteren yönlü döngüsüz grafik (Y.D.G) oluşturma işlemidir. Parametre öğrenme, ebeveynleri bilinen değişkenlerin koşullu olasılık değerlerinin tahmin edilmesidir.

Kısıtlama tabanlı algoritmalar koşullu bağımsızlık testlerinden yararlanarak düğümler arası ilişkileri en iyi şekilde temsil edebilecek bir yapı oluşturmayı amaçlar. Skor tabanlı algoritmalar çok değişkenli bir veri seti için nedensellik yapısının doğruluğunu gösteren skor fonksiyonunu maksimize edecek şekilde bir Bayesci ağ oluşturur. Karma algoritmalar kısıtlama tabanlı algoritmalarda kullanılan koşullu bağımsızlık testlerinden ve skor tabanlı algoritmaların skor ölçütlerinden yararlanarak Bayesci ağ yapısını oluşturmakta-

dır. Bu algoritmaların temelindeki mantık kısıtlama tabanlı algoritmalarından yararlanarak bir yönsüz grafik oluşturmak ve ardından skor tabanlı algoritmaları uygulayarak kenarların yönlerini belirlemektir.

3.UYGULAMA

Bu çalışmada özel bir bankanın müşterilerinin kredi risklerini incelemek üzere kayıt altına aldığı bir veri seti kullanılmıştır. Çalışmada kullanılan değişkenler şunlardır:

- Kredi durumu
- Hesap Dengesi
- Kredinin Amacı
- Tasarruflar
- Yaş
- Medeni durum
- Cinsiyet
- Daire tipi

Bu çalışmada skor tabanlı algoritmalar ve karma algoritmalar kullanılarak en uygun Bayesci ağ yapısı belirlenmiştir. Öğrenilen yapıyı kullanarak, parametre öğrenme işlemi uygulanmıştır. Parametre tahminleri, en çok olasılık tahmini yöntemi kullanılarak elde edilmiştir. Bu çalışmada kullanılan yapı öğrenme algoritmaları şunlardır:

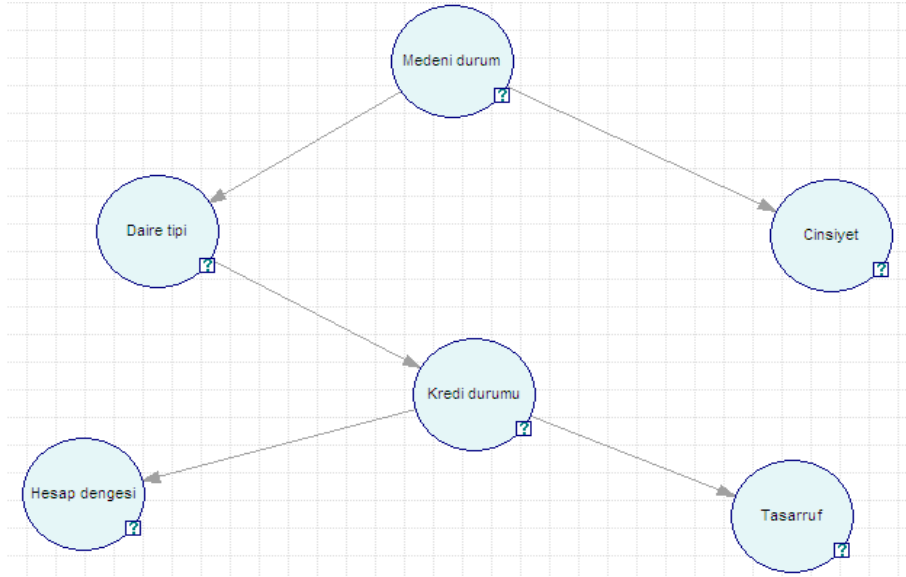
- Büyüme-daralma (GS)
- Artımsal ilişki (IAMB)
- Hızlı artımsal ilişki (FIAMB)
- Aralıklı artımsal ilişki (INTIAMB)
- En az en çok ebeveynler ve çocuklar (MMPC)
- Tepe tırmanma (HC)
- Tabu araştırma (TABU)

Yukarıda gösterilen yapı öğrenme algoritmaları kullanılarak karma yöntemler de uygulanmıştır. Elde edilen Bayesci ağların kalitesini ölçmek üzere Bayesci Bilgi Kriteri (BIC) skorları hesaplanmıştır.

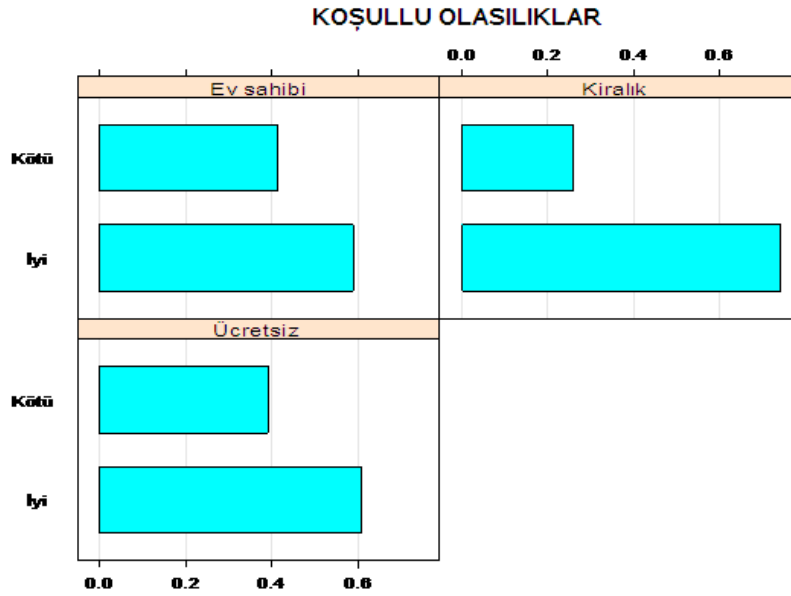
Tablo 1: Yapı öğrenme algoritmaları ile oluşturulan Bayesci ağların BIC skorları

ALGORİTMA	BIC SKORU
HC+GS	-8026,502
HC+IAMB	-7999,105
HC+INTER. IAMB	-7999,105
HC+FAST. IAMB	-8020,830
HC+MMPC	-7999,105
HC+GS	-8026,502
HC+IAMB	-7999,105
HC+INTER. IAMB	-7999,105
HC+FAST. IAMB	-8020,830
HC+MMPC	-7999,105
HC	-7996,949
TABU	-7996,949

Yukarıdaki tabloda kredi riski veri seti için oluşturulan Bayesci ağ yapılarının BIC skor değerleri gösterilmektedir. HC ve TABU skor tabanlı algoritmalar olup diğerleri karma algoritmalarıdır. Buna göre HC ve TABU algoritmaları ile oluşturulan yapılar daha yüksek skor değeri verdiği için, skor tabanlı algoritmaların karma algoritmalara göre daha yüksek performansa sahip oldukları sonucuna varılmaktadır. Tepe tırmanma ve tabu araştırma algoritmaları en iyi skor değerlerini verdiği için skor tabanlı algoritmalar kullanılarak Bayesci ağ yapısı oluşturulmuştur. Her iki algoritma da aynı skor değerine sahip olduğu için aynı Bayesci ağ yapılarına sahiptir. Koşullu olasılıkların hesaplanması için en çok olabilirlik metodu kullanılmıştır. Şekil 1’ de Bayesci ağ yapısına ilişkin grafik gösterilmektedir.

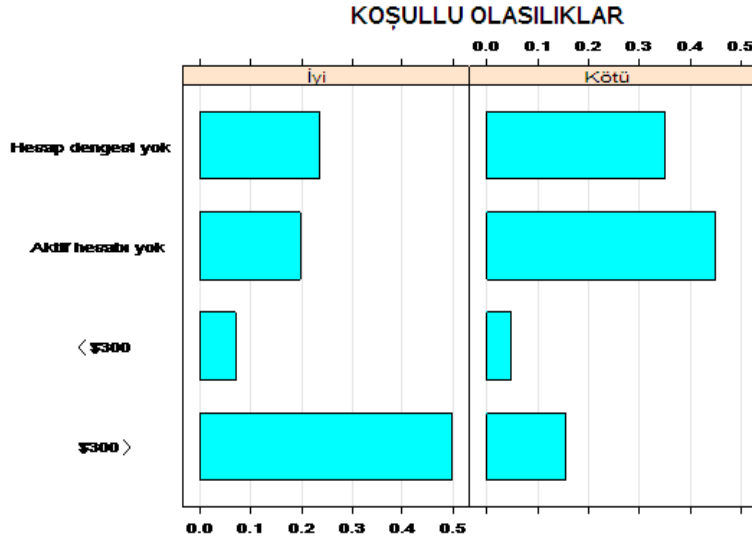


Şekil 1. Skor tabanlı algoritmalarla oluşturulan Bayesci ağ yapısı



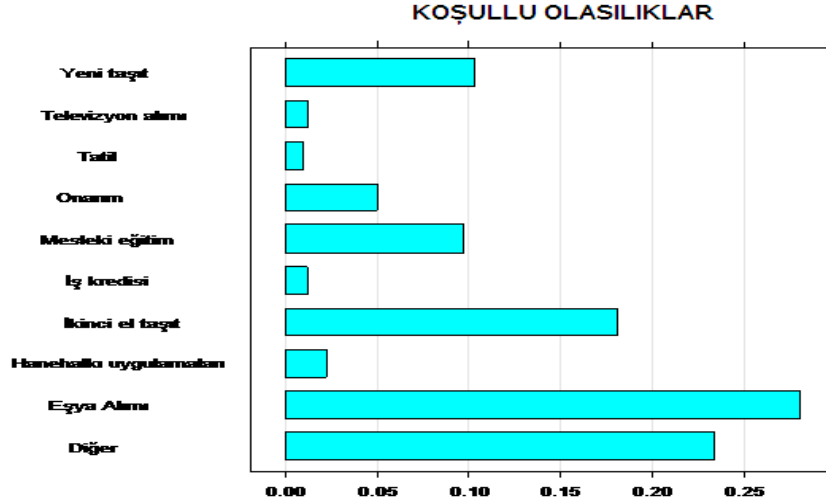
Şekil 2. Kredi durumu için koşullu olasılık grafiği

Kredi durumuna ilişkin mülkiyet durumu açısından tüm bireylerin kredi durumunun iyi düzeyde olduğu görülmektedir. Kiralık evde oturan bireylerin kredi durumu, diğer mülkiyet tiplerine sahip bireylere göre daha elverişlidir.



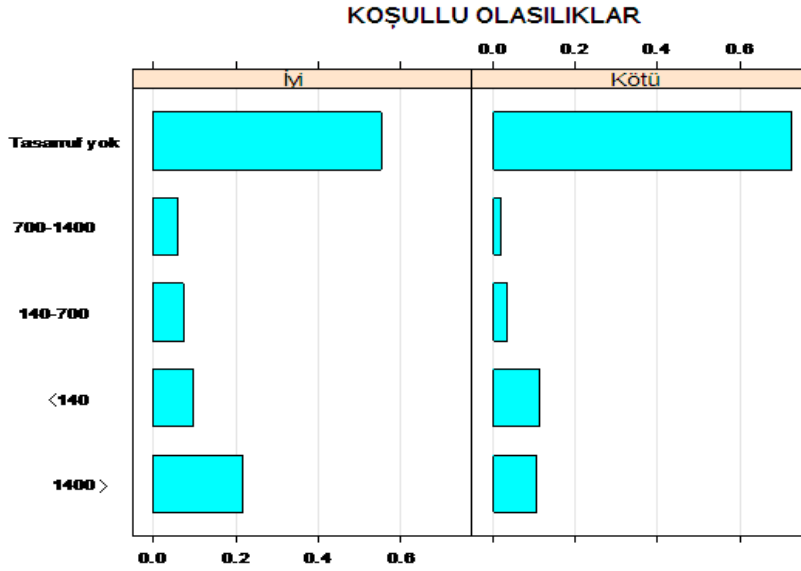
Şekil 3. Hesap dengesi için koşullu olasılık grafiği

Hesap dengesine göre 300 dolar ve üzeri aktif hesaba sahip olan bireylerin kredi durumunun iyi olduğu görülmektedir. Aktif hesabı olmayan bireylerin ise kredi durumu kötüdür.



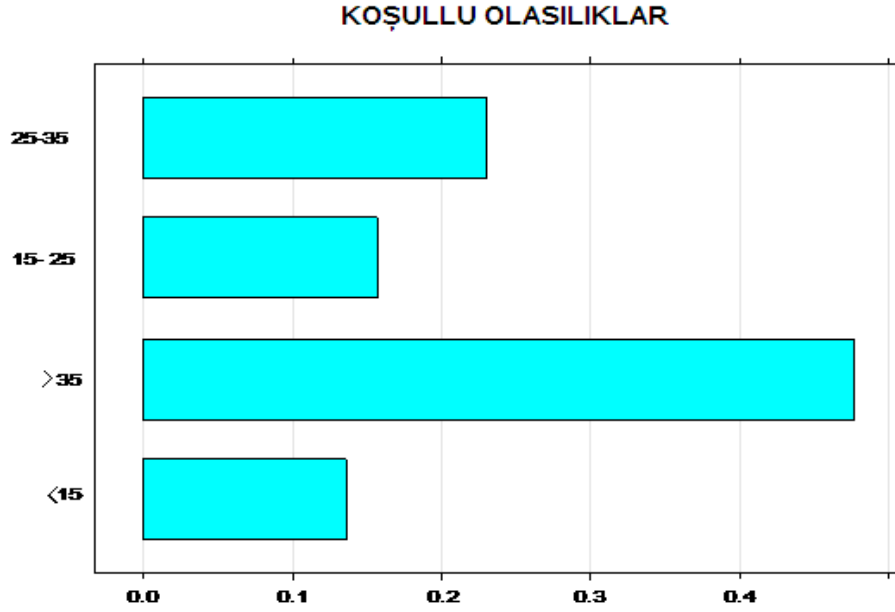
Şekil 4. Kredi amacı için koşullu olasılık grafiği

Kredi amacına ilişkin grafikte bireylerin alacakları kredileri hangi amaçla kullanacakları gösterilmektedir. Buna göre eşya alımı için kredi alacak olan bireylerin oranı en yüksek düzeyde iken, televizyon alımı, tatil ve iş kredisi için kredi alacak olan bireylerin oranı en düşük düzeydedir.



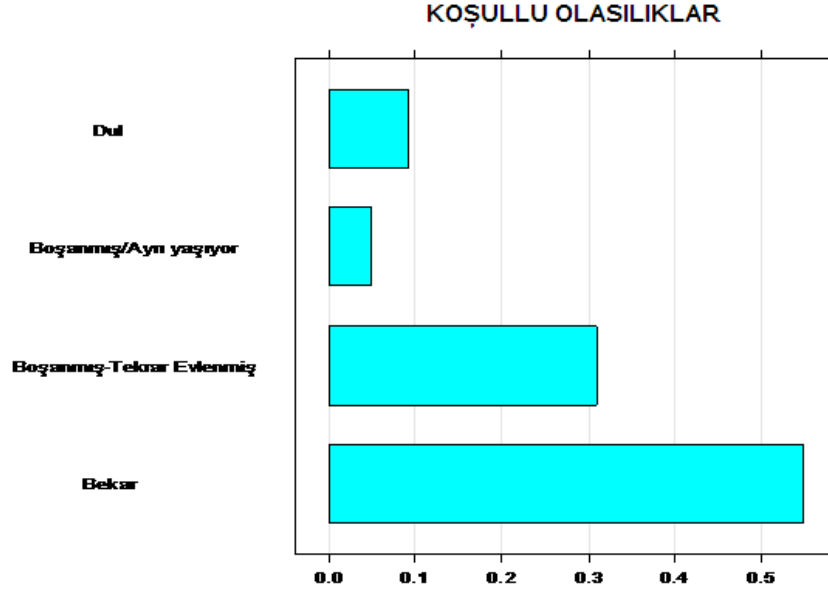
Şekil 5. Tasarruf için koşullu olasılık grafiği

Tasarruf durumuna ilişkin grafikte çoğunlukla kredi durumu iyi ve kötü olan bireylerin tasarruf u bulunmamaktadır.



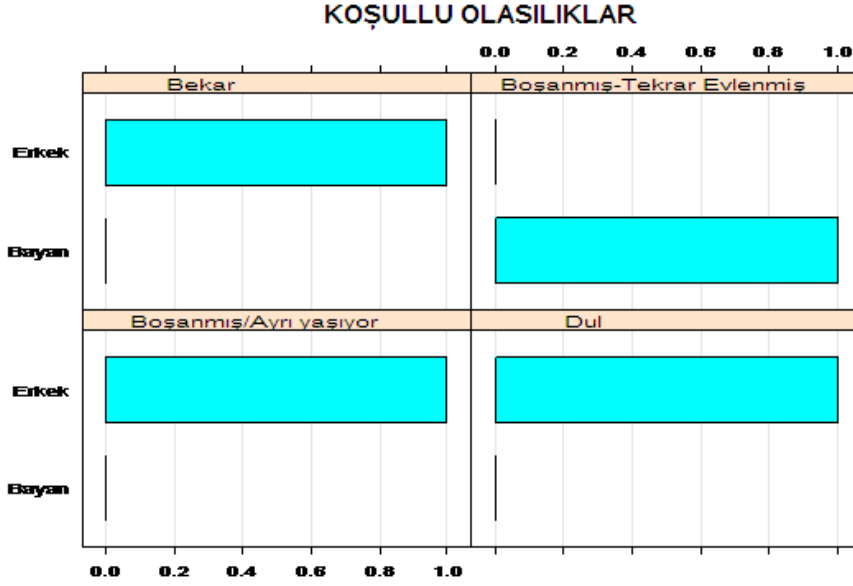
Şekil 6. Yaş için koşullu olasılık grafiği

Bireylerin yaş durumlarına ilişkin grafikte kredi kullanacak olan bireylerin çoğunluğunun 35 orta yaş ve üzeri olduğu görülmektedir.



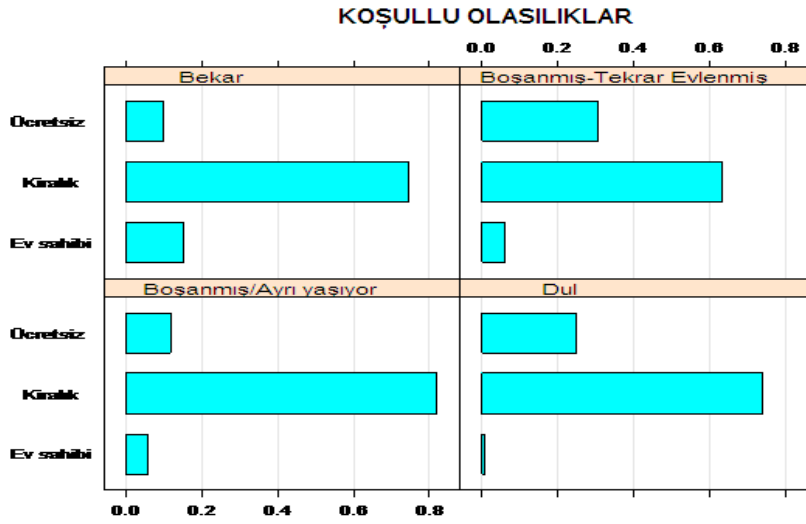
Şekil 7. Medeni durum için koşullu olasılık grafiği

Bireylerin yaş durumlarına ilişkin grafikte kredi kullanacak olan bireylerin çoğunluğunun bekar olduğu görülmektedir.



Şekil 8. Cinsiyet için koşullu olasılık grafiği

Bireylerin cinsiyetlerine ilişkin medeni durumları açısından erkeklerin bekar ve dul olduğu, bayanların ise boşanıp tekrar evlenmiş oldukları görülmektedir.



Şekil 9. Daire tipi için koşullu olasılık grafiği

Bireylerin ikamet ettikleri daire tiplerine ilişkin medeni durumları açısından tüm bireylerin çoğunlukla kiralık dairelerde ikamet ettikleri görülmektedir.

4.SONUÇ VE DEĞERLENDİRME

Bu çalışmada risk analizi alanında bankacılık sektörüne ilişkin risk verileri üzerinde Bayesci ağlar üzerine bir uygulama yapılmıştır. Uygulama sonucunda değişkenler arası ilişkileri en iyi temsil eden yapı öğrenme algoritması kullanılarak Bayesci ağ yapısı elde edilmiştir. Elde edilen skor değerlerine göre skor tabanlı yöntemlerin karma algoritmalara göre daha iyi sonuç verdiği belirlenmiştir. Skor tabanlı algoritmalar ile elde edilen Bayesci ağ üzerinden parametre öğrenme işlemi gerçekleştirilmiştir. Parametre öğrenme sonucunda her değişken için koşullu olasılık değerleri bar grafiği ile gösterilmiştir.

Bayesci ağ yapısı ile birlikte elde edilen olasılık değerlerine göre kiralık evde oturan ve aktif hesaba sahip olan bireylerin kredi durumlarının daha iyi olduğu belirlenmiştir. Bu çalışmada elde edilen Bayesci ağ yapısı ile birlikte oluşturulan model gelecekteki bireylerin kredi risklerinin hesaplanması için alternatif bir model olarak kullanılabilir. Sonuç olarak skor tabanlı algoritmalar kullanılarak oluşturulan Bayesci ağ yapıları ile risk analizi alanında daha doğru nedensellik ilişkileri ve tahmin değerleri elde edilmektedir. Gelecekteki çalışmalar için farklı yapı öğrenme algoritmaları ile birlikte daha doğru tahmin sonuçları elde edilebilir.

KAYNAKLAR

- Fenton N. E., Neil M. D., 2012. Risk Assessment and Decision Analysis with Bayesian Networks, CRC Press.
- Gemela J., 2001. Financial analysis using Bayesian Networks, Applied Stochastic Models in Business and Industry; 17:57-67.
- Neapolitan, R.E., Learning Bayesian Networks, 2003, Prentice Hall, Upper Saddle River, NJ.
- Margaritis, D., 2003. Learning Bayesian Network Model Structure from Data. PhD thesis, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA. CMU-CS 03-153.
- Oliva G. M., Weber P., Simon C., Iung B., 2009. Bayesian networks Applications on Dependability, Risk Analysis and Maintenance, 2nd IFAC Workshop on Dependable Control of Discrete System, DCDS'09, Bari : Italy .

-
- Scutari, M. 2011. Measures of Variability for Graphical Models. PhD thesis, Universita degli Studi di Padova, Dipartimento di Scienze Statistiche.
- Spirtes, P., Glymour, C., and Scheines, R., 2000. Causation, Prediction, and Search. IT Press.
- Tsamardinos, I., Aliferis, C., F., Statnikov, A., 2003. Algorithms for Large Scale Markov Blanket Discovery. In Proceedings of the 16th International Florida Artificial Intelligence Research Society Conference 376-381.
- Verma, T., S., Pearl, J. 1991. Equivalence and Synthesis of Causal Models. *Uncertainty in Artificial Intelligence*, 6, 255-268.
- Yaramakala, S., Margaritis, D., 2005. Speculative Markov Blanket Discovery for Optimal Feature Selection, In Proceedings of the 5th IEEE International Conference on Data Mining ,pages 809-812.

