

DOI: 10.53608/estudambilisim.\*\*\*\*\*\*\*

(Geliş Tarihi / Received Date: \*\*.\*\*.20\*\*, Kabul Tarihi/ Accepted Date: \*\*.\*\*.20\*\*)

| (**Araştırma Makalesi**) | |
| --- | --- |
| **Mahsül Sınıflandırmada Rastgele Orman ve Lojistik Regresyon Sınıflandırıcıları ile RFECV ve ANOVA Öznitelik Seçim Yöntemlerinin Değerlendirilmesi**  **Ad SOYAD\*1, Ad SOYAD1, Ad SOYAD2 (Bu alanda değişiklik yapmayınız. Kabul aşamasından sonra Ad Soyad yazılacaktır)**    1……….. Üniversitesi, ……….. Fakültesi, ………..Bölümü, 12345, Şehir, ORCID No : <http://orcid.org/>  2……….. Üniversitesi, ……….. Fakültesi, ………..Bölümü, 12345, Şehir, ORCID No : <http://orcid.org/> | |
|  | |
|  |  |
| **Anahtar Kelimeler:**  Mahsul türü sınıflandırması,  Makine öğrenmesi,  Akıllı tarım,  Mahsul haritalaması,  Optik ve radar verileri | **Özet:** Bu çalışmada, optik ve radar verileri kullanılarak elde edilen, toplamda 7 farklı ürün çeşidini içeren veri seti üzerinde, farklı makine öğrenmesi yöntemleri ile yürütülen bir sınıflandırma problemi üzerinde çalışılmıştır. Veri seti, Kanada’nın Winnipeg bölgesi üzerinden RapidEye ve UAVSAR uyduları aracılığıyla oluşturulmuştur. Çalışma kapsamında mahsullerin fiziksel ve biçimsel farklarının belirgin hale gelmesini sağlamak amacıyla polarimetrik ve optik öznitelikler kullanılmış, bitki örtüsü indeksi bilgisini içeren özniteliklerden de fayda sağlanmıştır. Öznitelik seçimi aşamasında, ANOVA ve Recursive Feature Elimination Cross Validation(RFECV) yöntemleri ile toplamda 174 öznitelik arasından farklı sayıda ve tipte öznitelik seçimi yapılmış, sınıflandırma aşamasına etkisi gözlemlenmiştir. Bunun yanında, mahsul sınıflandırması için kullanılacak olan Random Forest(RF) ve Logistic Regression gibi makine öğrenmesi yöntemlerinin karşılaştırmaları yapılmıştır. Sonuç olarak, en başarılı kombinasyonun tespiti sonucunda ortaya çıkan mahsul sınıflandırma aracı, tarım alanında doğal kaynakların doğru kullanımı, tarım arazisi monitörlemesi ve üretimdeki verimin artırılması gibi imkanlar ile akıllı teknolojilerin tarım sektörüne entegrasyonuna katkı sağlamaktadır. |
|  |  |
|  |  |
| (**Research Article**)  **Evaluation of Random Forest and Logistic Regression Classifiers with RFECV and ANOVA Feature Selection Methods in Crop Classification** | |
|  | |
|  | |
| **Keywords:**  Crop type classification,  Machine learning,  Smart agriculture,  Crop mapping,  Optical and radar data | **Abstract:** In this study, a classification problem carried out with different machine learning methods was studied on the data set containing 7 different product types in total, obtained using optical and radar data. The dataset was created via RapidEye and UAVSAR satellites over Winnipeg, Canada. Within the scope of the study, polarimetric and optical features were used to make the physical and formal differences of the crops more evident, and the features containing the vegetation index information also benefited. During the feature selection stage, different numbers and types of features were selected from 174 features using ANOVA and Recursive Feature Elimination Cross Validation (RFECV) methods, and its effect on the classification stage was observed. In addition, comparisons of machine learning methods such as Random Forest (RF) and Logistic Regression, which will be used for crop classification, were made. As a result, the crop classification tool, which emerged as a result of the determination of the most successful combination, of the integration of smart technologies into the agricultural sector with opportunities such as the correct use of natural resources in the field of agriculture, crop monitoring and increasing the production efficiency. |
|  |  |

# 1. GİRİŞ

Günümüzde, insan popülasyonun giderek artması ile gıda sektöründe mahsul teminatı mevcut sisteminin yetersizliği sonucu bazı problemlerin ortaya çıkabileceği düşünülmektedir. Son yıllarda teknolojideki hızlı gelişim, tarım sektörünü de kapsayacak çalışmaların ve yeniliklerin ortaya çıkmasını da beraberinde getirmiştir. Yapay zeka ve makine öğrenmesi yöntemleri kullanılarak tarım sektöründeki talebin karşılanmasını ve üretimde verimin artırılmasını hedefleyen, tarım alanlarının sınıflandırılması, bitkilerin monitörlenmesi ve mahsul haritalanması gibi alanlarda yapılmış çalışmalar bulunmaktadır. Tarım arazisi ve mahsul haritalanması, tarım arazilerinde yetiştirilen mahsullerin tespiti ve ayırt edilmesi için kullanılan bir yöntemdir. Bu yöntemin kullanımıyla birlikte maliyet optimizasyonu, doğal kaynakların verimli kullanılması ve akıllı uygulamalar ile üretimin iyileştirilmesi hedeflenmektedir. Genellikle uydu görüntüleri ve sensör verilerinin makine öğrenmesi ve yapay zeka teknolojileri kullanılarak işlenmesi sonucunda tarım arazilerinin mahsul tiplerine göre sınıflandırılması ve haritalanması yapılmaktadır.

Şimdiye kadar yapılan çalışmalar incelendiğinde ilk olarak; zaman, spektral, doku ve polarimetrik gözlemler bir araya getirilerek tarla haritalama performansının arttırılması amaçlanmıştır. Zaman içindeki değişiklikleri takip etmek için temporal gözlemler, bitki örtüsünün spektral özniteliklerini incelemek için spektral gözlemler, bitki dokusunu analiz etmek için doku gözlemler ve bitki örtüsünün polarimetre ile ölçülen özniteliklerini kullanmak için ise polarimetrik gözlemler seçilmiştir. öznitelik seçimi işleminde bu dört gözlemin birleştirilmesi kullanılmıştır. Sınıflandırma modeli olarak ise Random Forest tercih edilmiştir. Gözlemlerin birleştirilmesi yöntemiyle sınıflandırma işleminin doğruluğunda artış elde edilmiştir. Sonuç olarak %90 genel doğruluk oranı ve %88 kappa skoru elde edilmiştir [1].

Başka bir çalışmada ise, farklı kaynaklardan gelen veriler, farklı ölçeklere ve dağılımlara sahiptir. Bu farklı dağılımdaki verilerin uyumlu bir şekilde birleştirilerek kullanılması için ensemble yöntemi kullanılmıştır. Ensemble yöntemin ile döndürme kalibre edilmiş en küçük kareler destek vektör makinesi (RLS-SVM) kullanılarak veri sınıflandırılmasını gerçekleştirilmiştir. Çalışmada, belirtilen yöntemler kullanılarak farklı kaynaklardan elde edilen verileri birleştirerek daha iyi bir sınıflandırma performansı elde edilmeye çalışılmıştır.

Sonuç olarak genel doğruluk oranı %89 olarak hesaplanmıştır [2].

Bahsedilenlerden farklı olarak, başka bir çalışmada öznitelik seçim yöntemi değiştirilmiştir. Maksimum ayırt edicilik ve minimum bağımlılık öznitelik seçimi yöntemi olan MSMD kullanılmıştır. MSMD, öznitelik seçiminde maksimum ayırt edicilik ve minimum bağımlılık ilkesini kullanarak en uygun öznitelikleri belirler. Sınıflandırma performansını artırmak için en iyi ve bağımsız öznitelikleri seçmeyi hedeflemektedir. Çalışmada, MSMD’nin diğer öznitelik seçimi yöntemlerine kıyasla daha iyi performans sergilediği ve daha az bağımlılık içerdiği bulunmuştur. Sonuç olarak %90 genel doğruluk oranı ve %88 kappa skoru elde edilmiştir [3].

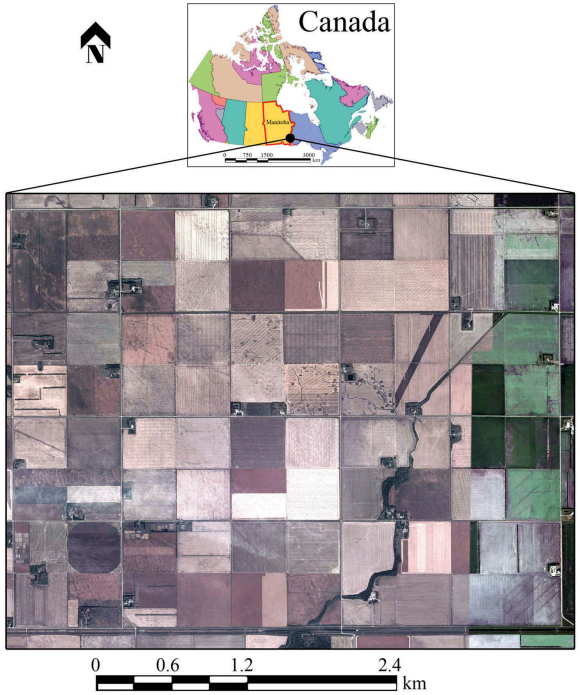
Şimdiye kadar yapılmış çalışmalar baz alındığında, tarım arazisi ve mahsul tipi sınıflandırması için polarimetrik, bitki örtüsü indeksi ve optik özniteliklerin ağırlıklı olarak kullanıldığı görülmüştür [4]. Bu kapsamda mahsul tiplerinin arasındaki yapısal ve biçimsel farkları daha belirgin hale getirecek özniteliklerin kullanımı da sonuçların iyileşmesini sağlamaktadır [5]. Yapılacak olan doğru sınıflandırmalar sonucunda bölgelere ait dominant mahsul tiplerinin anlaşılabilmesi ve tarım arazisi üzerindeki dağılımın uydu üzerinden takibi sağlanabilmektedir. Makine öğrenmesi yöntemleri bu takibi yapmak için kullanılmakta, tarımda planlama ve üretim süreçlerinin akılcı bir politika güdülerek yürütülmesine olanak tanımaktadır.

Bu çalışmada, tarım arazilerinin sınıflandırılması amacıyla, iki farklı zamana ait optik ve radar verilerinin birleştirildiği Winnipeg veri seti üzerinde farklı öznitelik seçimi ve sınıflandırma yöntemleri kullanılarak karşılaştırma yapılması planlanmaktadır. Öznitelik seçimi aşamasında Varyans Analizi(ANOVA) ve Recursive Feature Elimination Cross Validation(RFECV) yöntemleri kullanılarak en iyi sonuç alınan yöntemin seçilmesi hedeflenmektedir. Seçilen öznitelikler ile mahsul tipi sınıflandırması için ise Random Forest(RF) ve Logistic Regression sınıflandırıcı metotlarının kullanılması planlanmaktadır.

# 2. Materyal ve Metot

## 2.1. Veri seti

Bu çalışma kapsamında Kanada’nın Manitoba eyaletinde bulunan Winnipeg bölgesi temel alınarak 2012 yılında oluşturulmuş, optik ve radar verilerinden oluşan bir veri seti kullanılmıştır. Veri seti, kanola, mısır, soya fasulyesi, bezelye, yulaf, buğday ve geniş yapraklı bitkiler olmak üzere toplamda 7 farklı mahsul türünden oluşmaktadır. Optik veriler RapidEye uyduları, radar verileri ise İnsansız Hava Aracı Sentetik Açıklıklı Radar (UAVSAR) yardımıyla toplanmıştır. Veri seti, iki farklı tarihe ait 98’i radar ve 76’sı optik verilerden elde edilmiş olan toplamda 174 öznitelik içermektedir. Bu öznitelikler arasında, mahsullerin yapısal ve biçimsel özniteliklerini ayırt etme konusunda belirleyici olabilecek bitki örtüsü indeksleri, imge renk kanalları ve VV, VH gibi polarizasyon bilgisine sahip öznitelikler bulunmaktadır [3].



**Şekil 1.** Veri Seti Uydu Görüntüsü [1]

## 2.2. Öznitelik Seçimi

Öznitelik seçimi, makine öğrenmesi için gerekli bir adım olarak görülmekte ve veri setindeki en yoğun bilgiye sahip öznitelikleri seçmeyi amaçlamaktadır. Veri setlerinde çok sayıda öznitelik bulunabilir ve bazı öznitelikler diğerlerine göre daha fazla ayırt edici bilgi içermektedir. Gereksiz veya gürültülü özniteliklerin seçilmesi, modelin performansını olumsuz etkileyebilir ve gereksiz hesaplama maliyetine sebep olabilir. Ayrıca öznitelik seçimi, modelin sadece en anlamlı ve bilgilendirici öznitelikler üzerinde çalışmasını sağlaması sebebiyle modelin genel performansını artırmakta ve aşırı uyum (overfitting) riskini azaltmaktadır [6]. Bu çalışma kapsamında ANOVA ve RFECV öznitelik seçimi metotları kullanılarak deneyler yapılmış, en verimli sonucu veren yöntemin hesaplanması üzerinde çalışılmıştır.

ANOVA, veri gruplarının ortalamaları arasındaki varyans farkını test etmek ve bu farkın istatistiksel olarak anlamlı olup olmadığını belirlemek için kullanılan istatistiksel bir yöntemdir. Bu yöntem, özniteliklerin sınıflandırma performansına olan etkisini değerlendirmek amacıyla kullanılmaktadır. İlk adımda, her bir öznitelik sınıf etiketlerine göre gruplara ayrılmaktadır. Ardından, gruplar arasındaki varyansın grup içi varyansa göre anlamlı bir şekilde büyüklüğü test edilmektedir. Elde edilen olasılık değerleri(p), özniteliklerin sınıflandırmadaki önemini değerlendirmek için kullanılmaktadır.. Belirli bir anlamlılık düzeyinde (genellikle 0.05 olarak kabul edilir) p değerinden küçük olan öznitelikler, sınıflandırmadaki etkileri anlamlı kabul edilerek seçilmektedir [7]. Bu çalışma kapsamında, ANOVA yöntemi, anlamlı özniteliklerin seçilmesi aşamasında kullanılmıştır.

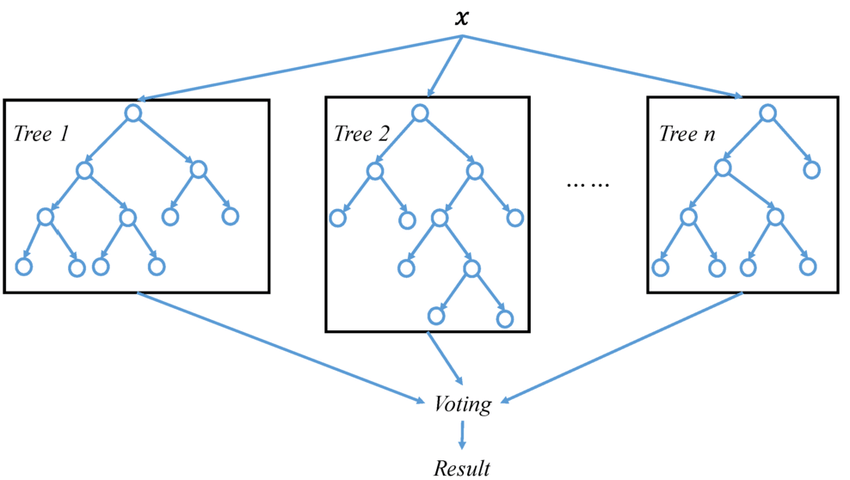
Bir başka öznitelik seçim tekniği olan RFECV ise geriye doğru eleme tekniği kullanılarak öznitelikler arasında karşılaştırmalı öznitelik seçimini sağlayan yöntemdir. Bu yöntemde, başlangıçta tüm özniteliklerle bir model oluşturulmaktadır. Ardından, en az öneme sahip öznitelikler çıkartılarak modelin performansı ölçülmektedir. Bu süreç tekrarlanarak her adımda en önemsiz öznitelik çıkarılmakta ve modelin performansı değerlendirilmektedir. Bu sayede, en önemli özniteliklerin belirlenmiş olduğu bir önem sıralaması elde edilmektedir [8]. Sonuç olarak, en verimli sonucu veren öznitelik sayısı ve sıralaması belirlenmektedir. Bu çalışmada, RFECV öznitelik seçimi yöntemi kullanılarak 174 öznitelik arasından model tarafından belirlenen sayıdaki en önemli olarak görülen özniteliklerin seçilmesi sağlanmaktadır.

## 2.3. Sınıflandırma

Sınıflandırma, makine öğrenmesinde farklı sınıflara ait verilerin ayırt edilmesi problemlerini çözmek için kullanılan bir tekniktir. Sınıflandırma, verilerin sahip olduğu belirli parametreleri kullanarak kategorilere (sınıflara) ayırmayı hedeflemektedir. Sınıflandırma algoritmaları, veri setindeki özniteliklere dayanarak bir örnek veya gözlemin hangi sınıfa ait olduğunu tahmin etmek için öğrenme süreci gerçekleştirmektedir. Bu süreçte, eğitim verileri kullanılarak bir model oluşturulmakta ve ardından bu model yeni, görülmemiş verileri sınıflandırmak için kullanılmaktadır. Sınıflandırma, denetimli öğrenme problemlerinde yaygın olarak kullanılan bir yaklaşımdır ve farklı algoritmalar, veri setinin özniteliklerine ve yapılandırmasına göre tercih edilebilmektedir. Bu çalışma kapsamında 7 farklı mahsul tipine ait verinin sınıflandırılması problemi için Random Forest(RF) ve Logistic Regression gibi makine öğrenmesi yöntemleri kullanılmaktadır.

Logistic Regression, bağımlı bir değişkenin iki veya daha fazla sınıfa ait olma olasılığını tahmin etmek için kullanılan bir sınıflandırma algoritmasıdır. Bu algoritma, veri setindeki öznitelikler ve sınıf etiketleri arasındaki ilişkiyi bir lojistik fonksiyon yardımıyla modellemeye çalışmaktadır. Lojistik fonksiyon, verilerin bir sınıfa ait olma olasılığını 0 ile 1 arasında bir değer olarak tahmin etmek için kullanılmaktadır. Model eğitimi sırasında maksimum olabilirlik yöntemi kullanılmakta ve modelin katsayıları tahmin edilmektedir [9].

Random Forest, birden fazla karar ağacının bir araya getirilmesi ile oluşturulan bir ensemble algoritmasıdır. Her bir karar ağacı, rastgele seçilen öznitelikler ve örnekler üzerinden bağımsız olarak eğitilmektedir. Ardından, bu ağaçların tahminleri bir araya getirilerek en iyi sonucu verecek ortak bir tahmin oluşturulmaktadır. Random Forest, genellikle özniteliklerin veri setindeki gürültüye dayanıklı olması ve iyi bir genelleştirme performansı sergilemesi açısından tercih edilmektedir [10].



**Şekil 2.** Random Forest İllüstrasyonu [11]

Her iki algoritmanın da kullanılmasının amacı, probleme farklı yaklaşımlarda bulunmak ve veri seti içerisindeki farklı özniteliklerin ön plana çıkarılması ile birlikte deneysel sonuçların incelenmesidir. Bu kapsamda Logistic Regression, sınıf etiketlerine yönelik doğrusal ilişkiyi değerlendirirken Random Forest daha karmaşık ilişkileri yakalamak ve doğrusal olmayan etkileri ele almak için daha uygundur. Bu şekilde, hem Logistic Regression, hem de Random Forest sınıflandırma yöntemlerinin performansını karşılaştırarak daha kapsamlı bir analiz yapma imkanı elde edilmiştir.

Sonuç olarak çalışma kapsamında kullanılacak metotlar, makine öğrenmesi temel adımlarından bazıları olarak görülen, veri seti içerisindeki önem arz eden özniteliklerin seçilmesinde ve verinin belirli kategorilerde sınıflandırılmasında önemli rol oynamaktadır. Bu adımlar, modelin performansını artırmakta, gereksiz öznitelikleri elemekte, önemli bilgileri vurgulamakta ve sınıflandırma problemlerini çözmek için güçlü algoritmaların kullanımını sağlamaktadır. ANOVA ve RFECV yöntemleri birlikte kullanılarak, hem istatistiksel anlamlılığı olan öznitelikler belirlenmekte, hem de karşılaştırmalı olarak en önemli özniteliklerin seçilmesi sağlanmaktadır. Bu şekilde, daha etkili bir öznitelik seti elde edilerek model performansının artırılması amaçlanmaktadır. Öznitelik seçiminin ardından, bu çalışmada sınıflandırma problemini çözmek için Logistic Regression ve Random Forest algoritmalarını kullanılmaktadır. Yapılan deneyler sonucunda farklı kombinasyonların karşılaştırılmakta, en başarılı sonucu veren öznitelikler ve sınıflandırma yöntemleri incelenmektedir.

# 3. Bulgular

Proje kapsamında iki farklı öznitelik seçimi ve sınıflandırma yöntemi temel alınarak birçok deney gerçekleştirilmiştir. Yapılan bu deneylerin sonucunda geliştirilen RFECV + RF yaklaşımında bir diğer metodoloji olan ANOVA + Logistic Regression’a kıyasla hem sınıf hem de ortalama bazında daha verimli sonuçlar elde edilmiştir. Gerçekleştirilen deneyler sonucunda RFECV öznitelik seçim yöntemi kullanılarak optimum öznitelik sayısı 37 olarak belirlenmiş, bir başka öznitelik seçim yöntemi olan ANOVA ile 50-80 aralığında farklı sayılarda öznitelik seçilerek deneyler yapılmış, bu deneyler sonucunda optimum değer 65 olarak hesaplanmıştır. Tablo 1’de farklı öznitelik sayılarına göre elde edilen sonuçlar görülmektedir. Tabloya göre öznitelik sayısının belirli bir sınırı aşmasından sonra karmaşıklığa sebep olduğu ve modelin performansını olumsuz derecede etkilediği görülmektedir.

**Tablo 1**. ANOVA Öznitelik Sayısına Göre Elde Edilen Doğruluk Değerleri.

| Öznitelik Sayısı | 50 | 60 | 65 | 80 |
| --- | --- | --- | --- | --- |
| Accuracy | 0.9795 | 0.9815 | 0.9817 | 0.8811 |

Belirlenen optimum öznitelik sayıları kadar özniteliğin kullanıldığı sınıflandırma aşamasında RF modeli Logistic Regression modeline göre daha iyi sonuçlar ortaya koymaktadır. Tablo 2’de görüldüğü üzere elde edilen sonuçlar doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve kappa skoru metrikleri bazında karşılaştırılmıştır. Yapılan bu karşılaştırma sonucunda RF sınıflandırma yönteminin Logistic Regression sınıflandırıcısından her bir metriğin ortalaması bazında yaklaşık olarak %1 daha iyi sonuç verdiği görülmektedir. Fakat, işlem süresi açısından Logistic Regression modeli RF modeline göre çok daha hızlı sonuç vermiştir.

**Tablo 2**. Model Sonuçlarının Kıyaslanması.

|  | Accuracy | Precision | Recall | Kappa Score |
| --- | --- | --- | --- | --- |
| Logistic Regression | 0.9817 | 0.9817 | 0.9817 | 0.9769 |
| RF | 0.9916 | 0.9916 | 0.9916 | 0.9893 |

Elde edilen ortalama sonuçların ardından model performanslarının mikro açıdan değerlendirilmesi amacıyla her bir sınıf için doğruluk değeri hesaplanmıştır. Ortaya çıkan sonuçlar Tablo 3’de görüldüğü üzere makro perspektifte de olduğu gibi neredeyse tüm alt sınıflarda RF modelinin daha iyi performans sergilediğini göstermektedir. Özellikle az miktarda veri bulunan geniş yapraklı bitkiler sınıfı için %5’e yakın bir fark gözlemlenmektedir.

**Tablo 3**. Sınıflar Özelinde Doğruluk Değerleri.

|  | Kanola | Mısır | Soya F. | Bezelye | Yulaf | Buğday | Geniş Y. |
| --- | --- | --- | --- | --- | --- | --- | --- |
| LR | 0.98 | 0.98 | 0.99 | 0.98 | 0.95 | 0.97 | 0.90 |
| RF | 0.99 | 0.99 | 0.99 | 0.99 | 0.97 | 0.98 | 0.95 |

Sonuç olarak yapılan deneyler kapsamında RF modeli Logistic Regression modeline göre daha az öznitelik kullanarak makro ve mikro açılardan daha duyarlı, kesin ve doğru sonuçlar üretmiştir. Ancak işlem süresi göz önüne alındığında öznitelik seçimi ve sınıflandırma aşamalarında RF modeli Logistic Regression modeline kıyasla 4 kat daha fazla zaman harcamaktadır. Bu sebeple kullanım senaryosuna göre iki modelden probleme daha uygun olanın kullanılması önerilmektedir.

# 4. Tartışma ve Sonuç

Literatürde Winnipeg veri seti ile yapılmış olan çalışmalar incelendiğinde farklı yöntemler kullanılarak elde edilmiş farklı sonuçlar bulunmaktadır. Yapılan bir çalışmada RF sınıflandırma yöntemi kullanılmış ve %90 oranında ortalama doğruluk değeri elde edilmiştir [1]. Bir başka çalışmada ise Least Squares Support Vector Machine (LS-SVM) metodu kullanılarak yaklaşık %89 oranında doğruluk hesaplanmıştır [2]. Son olarak bir başka çalışmada ise Maximum Separability and Minimum Dependency (MSMD) öznitelik seçim yöntemi ve SVM, RF, Naive Bayes sınıflandırma yöntemleri kullanılmış, sonuç olarak yaklaşık %90 doğruluk elde edilmiştir [3].

Proje kapsamında geliştirdiğimiz yaklaşımlar bu yöntemlerle karşılaştırıldığında önemli derecede bir iyileşme gözlemlenmektedir. Tablo 4’te görüldüğü üzere geliştirmiş olduğumuz RF tabanlı yaklaşım, doğruluk ve kappa skoru bakımından diğer tüm yöntemlere kıyasla daha iyi performans göstermektedir. Ayrıca proje kapsamında geliştirilen diğer bir metot olan Logistic Regression da literatürdeki çalışmalara göre daha iyi performans sergilemektedir.

**Tablo 4**. Geliştirilen Yöntemlerin Literatürde Yapılmış Çalışmalar ile Kıyaslanması.

|  | Accuracy | Kappa Score |
| --- | --- | --- |
| ANOVA + Logistic Regression (ours) | 0.9817 | 0.97 |
| RFECV + RF (ours) | 0.9916 | 0.98 |
| RF [1] | 0.9015 | 0.88 |
| LS-SVM [2] | 0.8973 | - |
| MSMD + RF [3] | 0.9011 | 0.88 |

Sonuç olarak bu çalışma ile RFECV ve ANOVA öznitelik seçme algoritmaları kullanılarak RF ve Logistic Regression modellerinin Winnipeg veri seti üzerindeki performansı değerlendirilmiştir. Elde edilen bulgular, geliştirilen yöntemlerin literatürlerdeki diğer yöntemlere kıyasla daha iyi sonuçlar verdiğini göstermektedir. Elde edilen sonuçlar ile RFECV ve ANOVA öznitelik seçme algoritmalarının Winnipeg veri seti üzerinde etkili olduğu ve sınıflandırma performansını önemli ölçüde iyileştirdiği anlaşılmaktadır.

# Kaynakça

1. Khosravi, I., & Alavipanah, S. K. (2019). A random forest-based framework for crop mapping using temporal, spectral, textural and polarimetric observations. *International Journal of Remote Sensing*, *40*(18), 7221-7251.
2. Khosravi, I., Razoumny, Y., Hatami Afkoueieh, J., & Alavipanah, S. K. (2021). An ensemble method based on rotation calibrated least squares support vector machine for multi-source data classification. *International Journal of Image and Data Fusion*, *12*(1), 48-63.
3. Khosravi, I., Safari, A., & Homayouni, S. (2018). MSMD: maximum separability and minimum dependency feature selection for cropland classification from optical and radar data. International Journal of Remote Sensing, 39(8), 2159-2176.
4. Yan, S., Yao, X., Zhu, D., Liu, D., Zhang, L., Yu, G., ... & Yun, W. (2021). Large-scale crop mapping from multi-source optical satellite imageries using machine learning with discrete grids. International Journal of Applied Earth Observation and Geoinformation, 103, 102485..
5. Sun, C., Bian, Y., Zhou, T., & Pan, J. (2019). Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. Sensors, 19(10), 2401.
6. Brownlee, J. (2020). How to Choose a Feature Selection Method For Machine Learning. MachineLearningMastery.com. <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>
7. Salman Pathan, M., Nag, A., Mohisn Pathan, M., & Dev, S. (2022). Analyzing the impact of feature selection on the accuracy of heart disease prediction. arXiv e-prints, arXiv-2206.
8. Mustaqim, A. Z., Adi, S., Pristyanto, Y., & Astuti, Y. (2021, June). The effect of recursive feature elimination with cross-validation (RFECV) feature selection algorithm toward classifier performance on credit card fraud detection. In 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST) (pp. 270-275). IEEE.
9. Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. The journal of educational research, 96(1), 3-14.
10. Yin, L., You, N., Zhang, G., Huang, J., & Dong, J. (2020). Optimizing feature selection of individual crop types for improved crop mapping. Remote Sensing, 12(1), 162.
11. Wang, Y., Pan, Z., Zheng, J., Qian, L., & Li, M. (2019). A hybrid ensemble method for pulsar candidate classification. Astrophysics and Space Science, 364, 1-13.