Bilgisayar Tabanlı Optimizasyon Dersi

DOÇ. DR ÖMÜR TOSUN

Büşra Boyacı

Amaç

3 farklı makine öğrenme modeli kullanmak. Seçilen her model için en iyi sonucu alabilmek için öznitelik seçme ve parametre optimizasyonu yöntemlerini kullanmak.

Problemin Tanımı

Deniz sevk santrallerinin bakım veri kümesini kullandık. GT Kompresör Bozulma Durumu katsayısı

Çıktı 1'i GT Tribün bozulma durumu katsayısını çıktı2 'yi temsil ediyor. Bu çıktılardan tercih ettiğimiz birinin tahmini gerçekleştirecek bir model oluşturulacak. Tahmin edilecek olarak seçilen kolon "**cikti2**"dir.

Verilerin Yüklenmesi

File Reader'ı kullanarak veriler local bilgisayardan knime çalışma ortamına taşındı.

Verilerin İncelenmesi ve Ön İşlenmesi

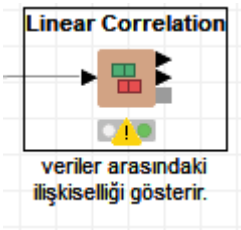
Verilerin genel görünümü aşağıdaki gibidir.

File Table - 0:1 - File Reader

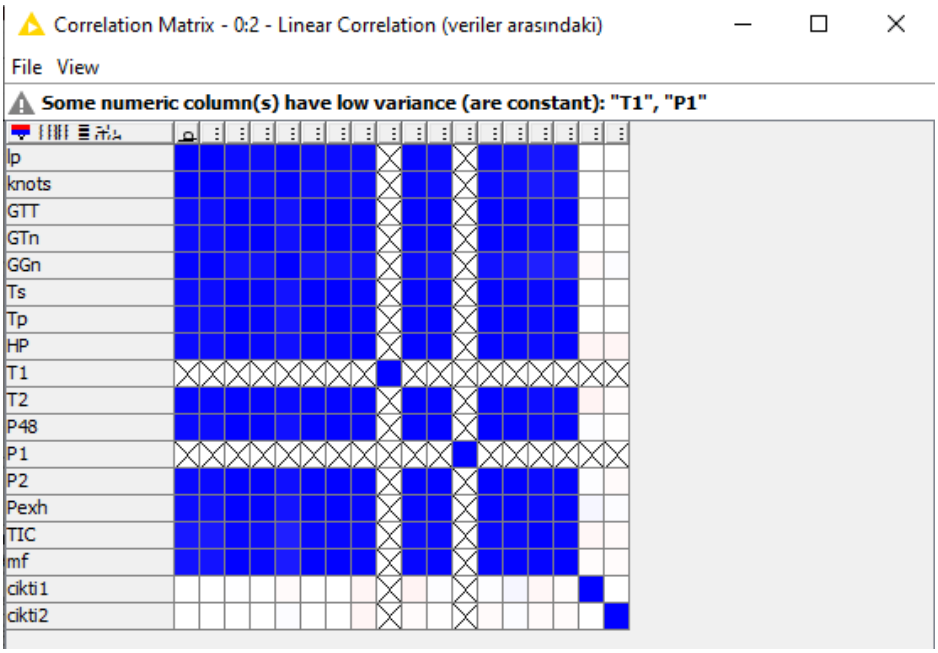
File Hilite Navigation View

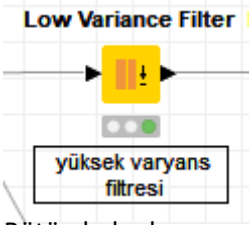
Table "grup2.txt" - Rows: 11934 Spec - Columns: 18 Properties Flow Variables

Row ID	lp	knots	GTT	GTn	GGn	Ts	Tp	HP	T1	T2	P48	P1	P2	Pexh	TIC	mf	cikti
Row0	1.138	3	289.964	1,349.489	6,677.38	7.584	7.584	464.006	288	550.563	1.096	0.998	5.947	1.019	7.137	0.082	0.95
Row1	2.088	6	6,960.18	1,376.166	6,828.469	28.204	28.204	635.401	288	581.658	1.331	0.998	7.282	1.019	10.655	0.287	0.95
Row2	3.144	9	8,379.229	1,386.757	7,111.811	60.358	60.358	606.002	288	587.587	1.389	0.998	7.574	1.02	13.086	0.259	0.95
Row3	4.161	12	14,724.395	1,547.465	7,792.63	113.774	113.774	661.471	288	613.851	1.658	0.998	9.007	1.022	18.109	0.358	0.95
Row4	5.14	15	21,636.432	1,924.313	8,494.777	175.306	175.306	731.494	288	645.642	2.078	0.998	11.197	1.026	26.373	0.522	0.95
Row5	6.175	18	29,792.731	2,307.404	8,828.36	246.278	246.278	800.434	288	676.397	2.501	0.998	13.356	1.03	35.76	0.708	0.95
Row6	7.148	21	38,982.18	2,678.086	9,132.429	332.077	332.077	854.747	288	699.954	2.963	0.998	15.679	1.035	45.881	0.908	0.95
Row7	8.206	24	50,996.808	3,087.561	9,318.562	437.989	437.989	952.122	288	741.77	3.576	0.998	18.632	1.04	62.44	1.236	0.95
Row8	9.3	27	72,763.329	3,560.395	9,778.528	644.905	644.905	1,115.797	288	789.094	4.498	0.998	22.811	1.049	92.556	1.832	0.95
Row9	1.138	3	379.88	1,355.375	6,683.916	7.915	7.915	464.017	288	550.985	1.1	0.998	5.963	1.019	3.879	0.079	0.95
Row10	2.088	6	6,969.176	1,371.94	6,828.438	27.424	27.424	635.96	288	581.44	1.33	0.998	7.272	1.019	12.785	0.289	0.95
Row11	3.144	9	8,379.307	1,386.758	7,114.396	60.353	60.353	605.169	288	587.437	1.389	0.998	7.567	1.02	13.052	0.258	0.95
Row12	4.161	12	14,724.544	1,547.466	7,794.646	113.795	113.795	660.568	288	613.688	1.657	0.998	8.998	1.022	18.066	0.358	0.95
Row13	5.14	15	21,636.777	1,924.313	8,495.697	175.314	175.314	730.495	288	645.457	2.078	0.998	11.185	1.026	26.316	0.521	0.95
Row14	6.175	18	29,792.983	2,307.388	8,829.394	246.295	246.295	799.298	288	676.17	2.501	0.998	13.342	1.03	35.687	0.707	0.95
Row15	7.148	21	38,981.896	2,678.086	9,132.932	332.208	332.208	853.701	288	699.779	2.963	0.998	15.661	1.035	45.799	0.907	0.95
Row16	8.206	24	50,997.234	3,087.583	9,318.935	438.098	438.098	951.073	288	741.605	3.575	0.998	18.611	1.04	62.342	1.234	0.95
Row17	9.3	27	72,763.515	3,560.401	9,779.311	644.963	644.963	1,114.887	288	788.949	4.496	0.998	22.784	1.049	92.448	1.83	0.95
Row18	1.138	3	753.021	1,371.886	6,697.838	8.858	8.858	473.665	288	553.143	1.117	0.998	6.047	1.019	0	0.088	0.95
Row19	2.088	6	6,961.726	1,371.057	6,830.364	27.259	27.259	635.133	288	581.244	1.33	0.998	7.262	1.019	13.192	0.289	0.95
Row20	3.144	9	8,379.456	1,386.759	7,116.987	60.356	60.356	604.338	288	587.286	1.389	0.998	7.559	1.02	13.018	0.258	0.95
Row21	4.161	12	14,724.616	1,547.466	7,796.657	113.801	113.801	659.666	288	613.524	1.657	0.998	8.989	1.022	18.023	0.357	0.95
Row22	5.14	15	21,636.925	1,924.315	8,496.61	175.282	175.282	729.497	288	645.272	2.078	0.998	11.174	1.026	26.26	0.52	0.95
Row23	6.175	18	29,793.328	2,307.377	8,830.433	246.22	246.22	798.166	288	675.943	2.501	0.998	13.328	1.03	35.613	0.705	0.95
Row24	7.148	21	38,981.432	2,678.086	9,133.432	332.108	332.108	852.658	288	699.604	2.962	0.998	15.644	1.035	45.717	0.905	0.95
Row25	8.206	24	50,997.566	3,087.601	9,319.306	438.134	438.134	950.027	288	741.441	3.574	0.998	18.589	1.041	62.245	1.232	0.95
Row26	9.3	27	72,764.565	3,560.403	9,780.103	645.076	645.076	1,113.985	288	788.806	4.495	0.998	22.756	1.049	92.342	1.828	0.95
Row27	1.138	3	1,341.723	1,391.603	6,714.114	10.018	10.018	490.661	288	556.617	1.142	0.998	6.185	1.019	0	0.107	0.95
Row28	2.088	6	6,895.321	1,366.843	6,830.317	26.481	26.481	633.651	288	580.761	1.326	0.998	7.239	1.019	15.319	0.289	0.95
Row29	3.144	9	8,379.548	1,386.759	7,119.575	60.351	60.351	603.509	288	587.136	1.389	0.998	7.552	1.02	12.983	0.257	0.95
Row30	4.161	12	14,724.731	1,547.467	7,798.673	113.797	113.797	658.766	288	613.361	1.657	0.998	8.98	1.022	17.98	0.356	0.95
Row31	5.14	15	21,636.973	1,924.318	8,497.521	175.3	175.3	728.501	288	645.088	2.078	0.998	11.162	1.026	26.203	0.519	0.95
Row32	6.175	18	29,793.658	2,307.365	8,831.471	246.288	246.288	797.037	288	675.718	2.5	0.998	13.315	1.03	35.54	0.704	0.95
Row33	7.148	21	38,981.087	2,678.087	9,133.933	332.197	332.197	851.618	288	699.429	2.961	0.998	15.627	1.035	45.636	0.904	0.95
Row34	8.206	24	50,997.769	3,087.617	9,319.674	438.052	438.052	948.984	288	741.277	3.574	0.998	18.568	1.041	62.148	1.23	0.95
Row35	9.3	27	72,765.436	3,560.405	9,780.887	645.04	645.04	1,113.084	288	788.664	4.494	0.998	22.729	1.049	92.235	1.826	0.95
Row36	1.138	3	1,836.145	1,403.714	6,725.143	10.747	10.747	505.147	288	559.3	1.163	0.998	6.293	1.019	0	0.124	0.95
Row37	2.088	6	6,846.147	1,364.27	6,831.294	26.018	26.018	632.165	288	580.39	1.324	0.998	7.222	1.019	16.581	0.288	0.95
Row38	3.144	9	8,379.616	1,386.76	7,122.162	60.358	60.358	602.68	288	586.986	1.388	0.998	7.544	1.02	12.949	0.256	0.95



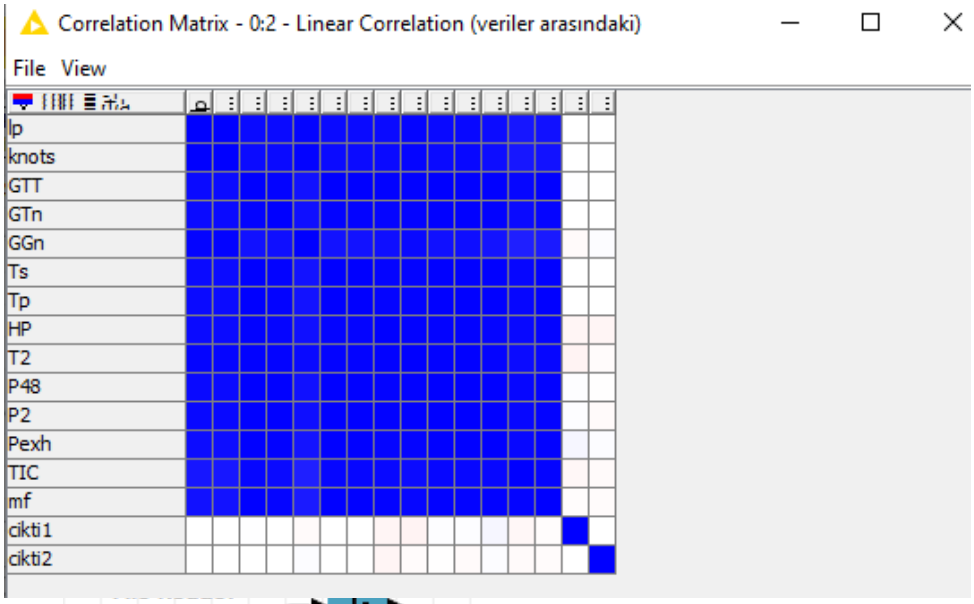
Veri setindeki kolonların birbiri arasındaki + yönlü veya – yönlü ilişkiyi gösteren Node'dur. File Reader'ı Linear Correlation Node'una bağladık. Görselleştirdiğimiz kolonlar arası ilişkiler aşağıdaki gibidir.



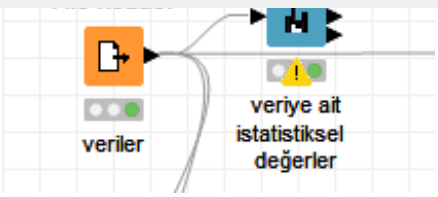


Low variance uyarısından sonra bu filtreyi kullanarak aşağıdaki grafik elde edildi.

Bütün kolonların aşırı koyu mavi ve aynı tonda olması kolonların çok ilişkili olduğu anlamına gelebilir.



Bu sebepten modele aldığımız değişkenlerde seçici olsak da skor değerlerinin yüksek olması beklenebilir.



Verilerin istatistik değerlerini incelediğimiz tablo aşağıdaki gibidir. Bu tablo Veri setinde eksik veri veya gürültülü veri denilen yanlışlıkla çok yüksek veya çok düşük değerler girilmiş gözlemler var mı incelememizi sağlar.

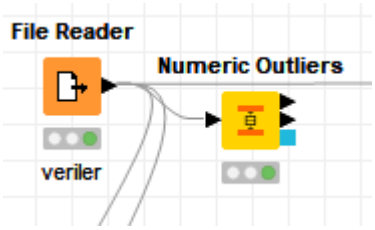
Statistics Table - 0:4 - Statistics (veriye ait)

File Hilite Navigation View

Table "default" - Rows: 18 | Spec - Columns: 16 | Properties | Flow Variables

Row ID	Column	Min	Max	Mean	Std. de...	Variance	Skewness	Kurtosis	Overall ...	No. mis...	No. NaNs	No. +∞s	No. -∞s	Median	Row co...	Histogram
lp	lp	1.138	9.3	5.167	2.626	6.898	0.022	-1.214	61,659	0	0	0	0	?	11934	1
knots	knots	3	27	15	7.746	60.005	0	-1.23	179,010	0	0	0	0	?	11934	3
GTT	GTT	253.547	72,784.872	27,247.499	22,148.613	490,561,06...	0.765	-0.509	325,171,64...	0	0	0	0	?	11934	254
GTn	GTn	1,307.675	3,560.741	2,136.289	774.084	599,205.855	0.567	-1.093	25,494,475....	0	0	0	0	?	11934	1.308
GGn	GGn	6,589.002	9,797.103	8,200.947	1,091.316	1,190,969.536	-0.14	-1.49	97,870,105....	0	0	0	0	?	11934	6.589
Ts	Ts	5.304	645.249	227.336	200.496	40,198.602	0.807	-0.43	2,713,025.058	0	0	0	0	?	11934	5

Bu tablodan çıkardığımız en önemli sonuç veriler arasında eksik gözlem olmadığıdır. Ayrıca Bu tablodan işimize yarayacağını düşündüğümüz bütün istatistiksal değerleri her kolon için gözlemlemek mümkün. Gözlemler arasındaki en büyük ve en küçük değeri birlikte görebilmek vb gibi.



Veri seti hakkında kolonlardaki veriler en çok tekrar eden vb gibi verileri sayısal olarak açıklamaya çalışan Node'dur. Görseli ve açıklaması aşağıdaki gibidir.

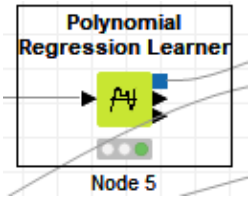
Summary - 0:3 - Numeric Outliers

File Hilite Navigation View

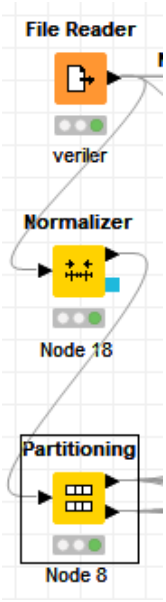
Table "default" - Rows: 18 | Spec - Columns: 5 | Properties | Flow Variables

Row ID	Outlier column	Member count	Outlier count	Lower bound	Upper bound
Row0	lp	11934	0	-2.862	13.154
Row1	knots	11934	0	-9	39
Row2	GTT	11934	0	-37,562.433	84,939.743
Row3	GTn	11934	0	-550.224	4,615.061
Row4	GGn	11934	0	3,946.83	12,244.078
Row5	Ts	11934	0	-347.755	740.437
Row6	Tp	11934	0	-347.755	740.437
Row7	HP	11934	0	223.543	1,200.419
Row8	T1	11934	0	288	288
Row9	T2	11934	0	404.331	867.686
Row10	P48	11934	0	-0.999	5.369
Row11	P1	11934	0	0.998	0.998
Row12	P2	11934	0	-4.869	27.974
Row13	Pexh	11934	0	0.996	1.06
Row14	TIC	11934	188	-32.654	90.877
Row15	mf	11934	0	-0.708	1.836
Row16	cikti1	11934	0	0.923	1.027
Row17	cikti2	11934	0	0.962	1.014

Aykırı olanlar başlığı altında bütün kolonlar bulunmaktadır. Bu kolonların özelliklerine bakıldığında; Her bir kolon için 11934 gözlem olduğunu görüyoruz. Bu gözlemler arasında aykırı değer sadece TIC



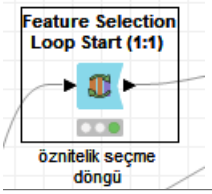
değeri için”188”dir. Diğer bütün kolonlar için aykırı olan değer “0”dır. Diğer sütunlarda her kolon için alt ve üst sınırları görmekteyiz.
Verilerilerin incelenmesi tamamlandı. Makine öğrenmesi modeli aşamasına geçilebilir.



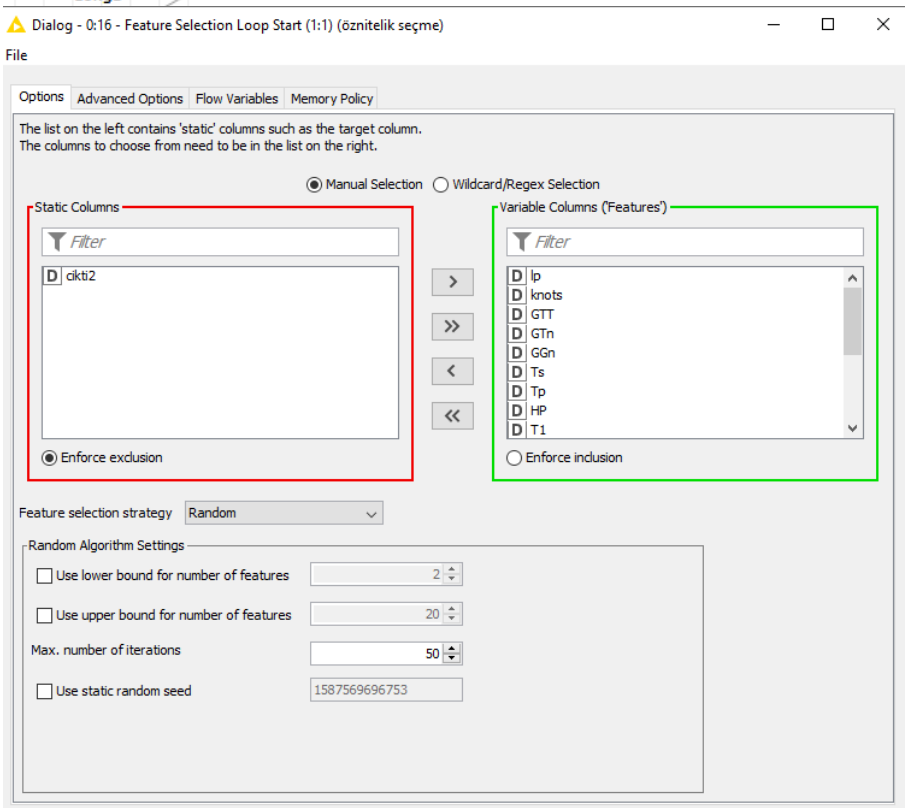
Verilerimiz sayısal ve sürekli değerlerden oluşmaktadır. Normalize etme işlemi yapılarak veriler 0 ile 1 aralığındaki değerlere dönüştürüldü. Modelin eğitimi için %70 veri eğitim, %30 veri test amacıyla ayrıldı.

Polynomial Regression Makine Öğrenmesi Modeli

Feature Selection



Kullanılacak makine öğrenmesi sırasında çalışıp en iyi öznitelikleri modelimize dahil etmemizi sağlayacak Node’dur.

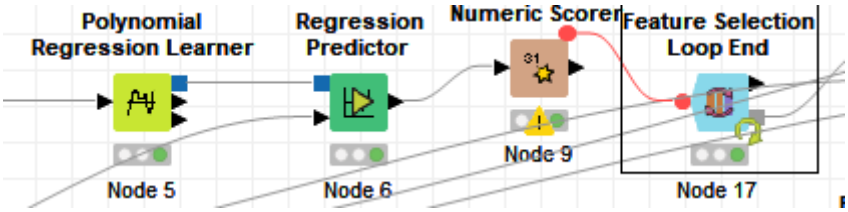


Veri setindeki tahmin kolonumuz olan ”cikti2” kolonu hariç bütün kolonları döngüye eklendi.

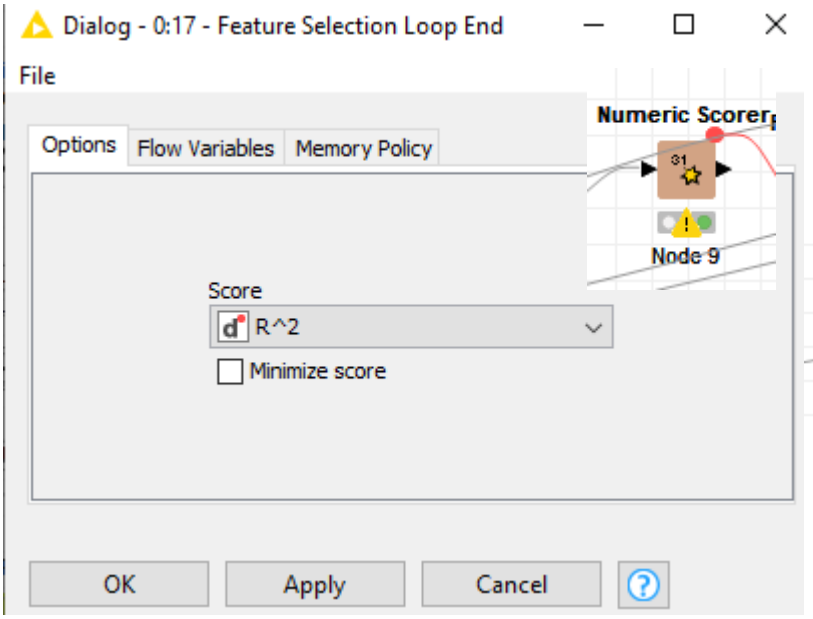
Feature Selection Node’u Polynomial

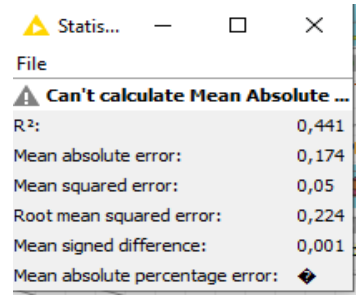
Regression Modeli'nin Learner Node'una bağlandı. Modelin özelliklerinde bir değiştirilme yapılmadı.

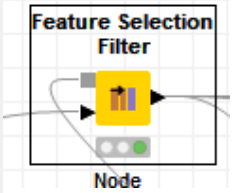
Eğitilmiş model ve test verileri Regression Predictor'a bağlandı. Tahmin sonuçları Numeric Scorer'a bağlandı. Feature Selection Loop End için R^2'yi maximum yapan özniteliklerin seçilmesi söylendi. Scorer'dan çıkan akış okları Feature Selection Loop End'e bağlandı ve özniteliklerin seçimi döngüsü sona erdirildi.



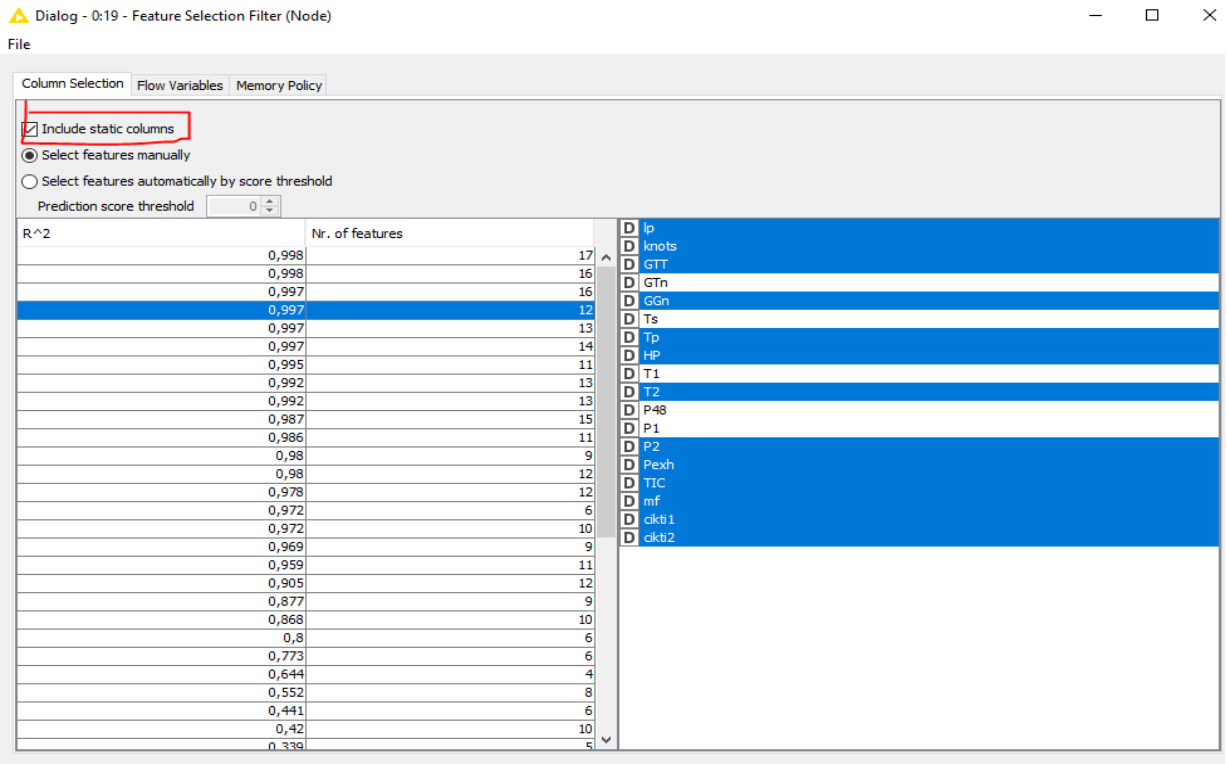
Döngü bitimindeki test verisi performansı skoru aşağıda gösterilmiştir.



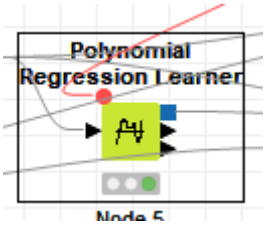




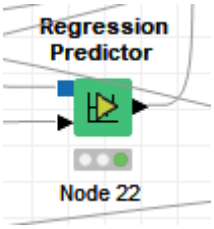
Döngü sonunda modele birlikte sokulması gereken özniteliklerin hepsini Feature Selection Filter'la görüntülendi. Node'un ayarları aşağıdaki görselde açıklanmıştır.



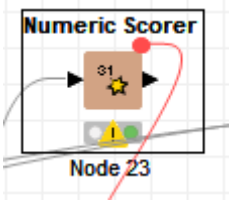
Kırmızıyla çizilen yer Feature Selection'a dahiletmediğimiz cikti2 değerini yeni kombinlenmiş olan değişkenlere otamatik dahil edilmesi sağlar. R^2 oranı 0,997 olan 12 değişkenin kombinasyonlandığı maviyle gösterilmiş öznitelikler seçildi. Modelden **"GTn, Ts, T1, P48, P1"** öznitelikleri çıkartıldı.



Feature Selection Filter'dan çıkan seçili öznitelikler Polynomial Regression Learner'a bağlandı. Bu modeli ikinci defa kullanıyoruz fakat bu kez ilk seferki gibi bütün değişkenler modele dahil edilmedi. Seçilen özniteliklerden başarılı olanlar kullanıldı.

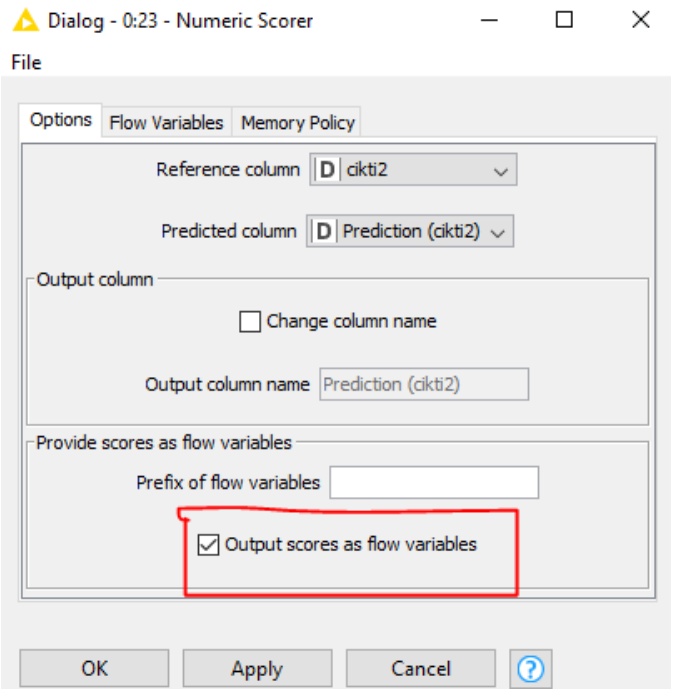


Polynomial Regression Learner mavi kutu çıkışından Predictor'un mavi kutu girişine bağlandı. Partitioning test kutu çıkışı Polynomial Regression Predictor'un siyah kutu girişine bağlandı.



Polynomial Regression Predictor çıkışından Numeric Scorer'a bağlandı.

Skorlama için karşılaştırılacak iki değer seçildi. Reference Column veri setindeki **“gerçek cikti2”** değerleri ile Predicted Column **“modelin tahminini temsil eden cikti2”** değerleri seçildi.

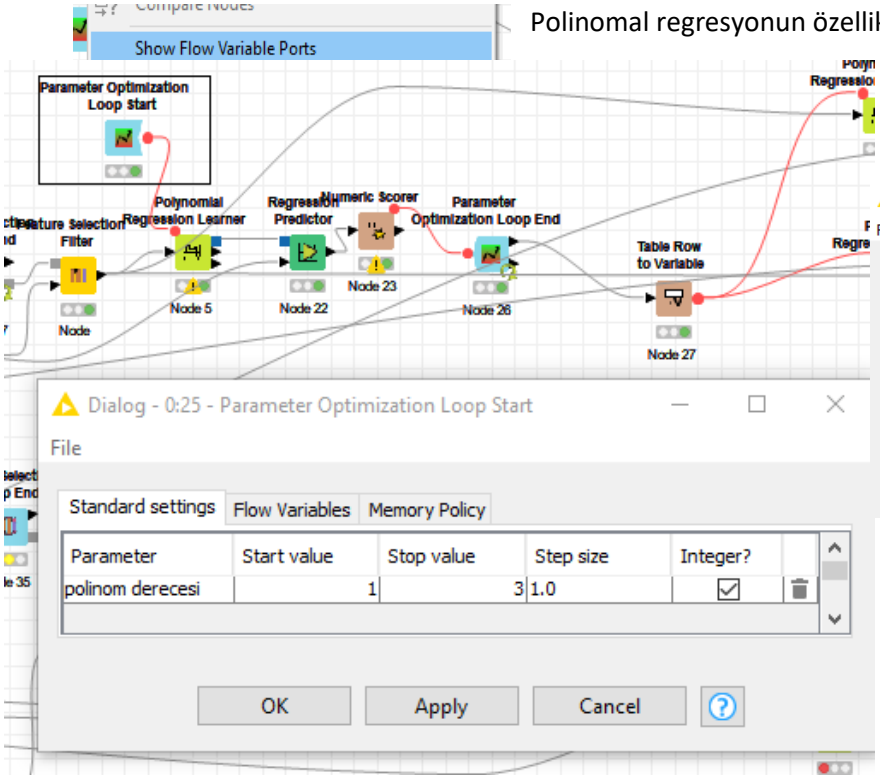


Kırmızıyla seçili bölge parametre optimizasyonu sırasında Numeric Scorer’dan Parametre döngüsüne veri akışı olmasını sağlar.

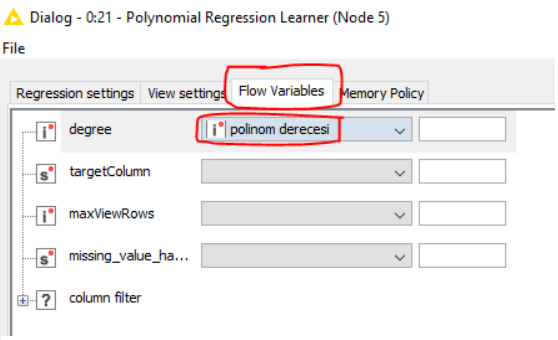
Parametre Optimizasyonu

En uygun öznitelikleri seçtikten sonra en uygun parametre değerleri için Parameter Optimization Loop Start Polynomial Regression Learner’a bağlandı.

Bağlantı için Polynomial Regression Learner’ın üzerine sağ tıklanarak akışı sağlayan özellik aktif edildi.

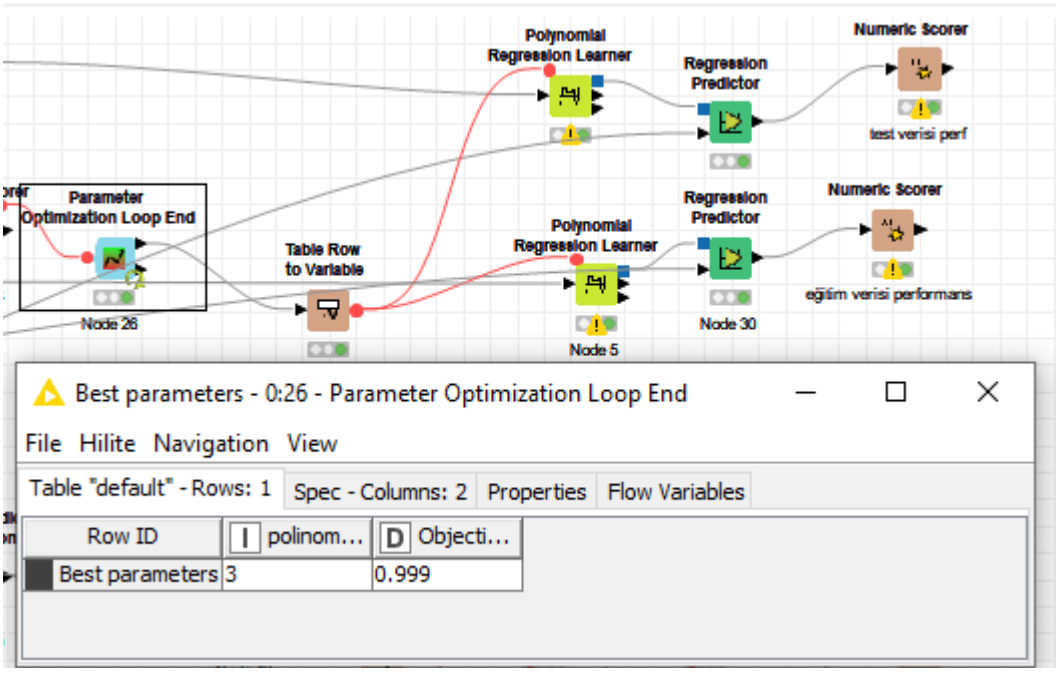


Polinomal regresyonun özelliklerine baktığımızda parametre olarak sadece **polinom derecesinin** kullanıldığını görebilmekteyiz.

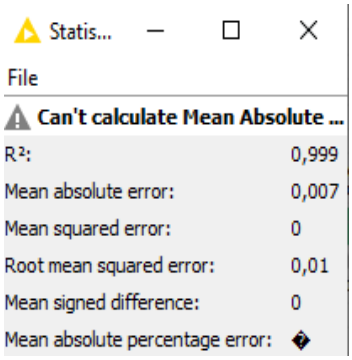


Parametre değerimizi belirledikten sonra kırmızı akış oklarını kullanarak Learner Node’una bu parametreleri aktarıyoruz.

Bizde parametre optimizasyonumuz için modelimize 1'inci ve 3'üncü dereceden polinomları modelde denenmesi için parametre optimizasyon döngüsünü başlatıldı.

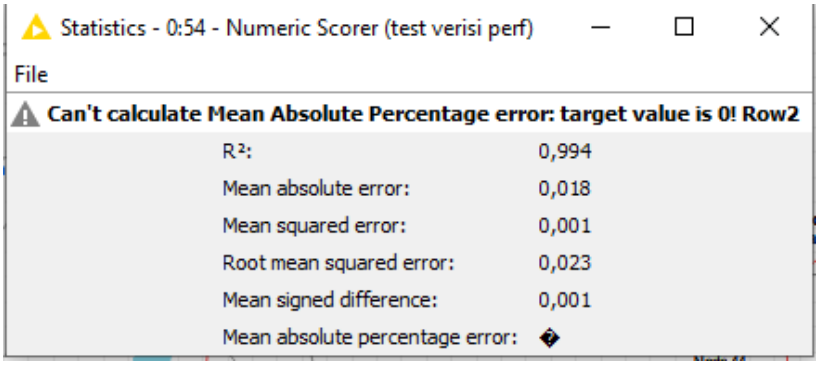


Döngü sonundaki scorer



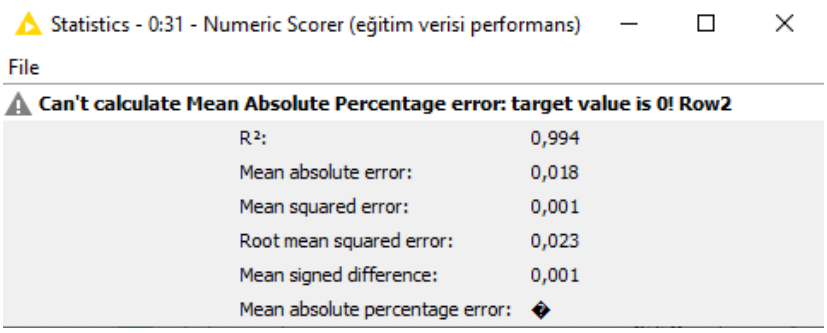
görüntüsü aşağıdadır. Döngü sırasında seçilen en iyi parametreler Table Row to Variable Nodu'ya modele aktarılabilecek biçime getirildi. Daha önce de yaptığımız **Flow Variable** ayarları yapıldı. Son olarak modelimize en iyi parametreleri de dahil ederek bir test ve eğitim performansı elde ettik.

Polynomial Regression Modeli Test Performansı



Regresyon modelleri için bir başarı katsayısı olan R^2 0,994 değerindedir. Bu modelin öğrendiği eğitim verilerinden bağımsız olarak hiç karşılaşmadığı verilerde gösterdiği performansı temsil eder.

Polynomial Regression Modeli Eğitim Performansı



Bu modelin eğitim sırasında öğrendiği verileri tekrar tahmin ederken gösterdiği performansı temsil eder.

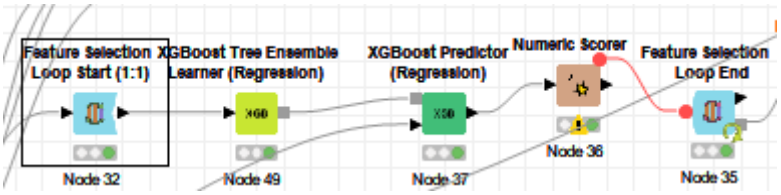
Sonuç:

Eğitim ve test performanslarının birbirine yakınlığı modelin öğrenme aşamasının başarılı olduğunu göstermektedir.

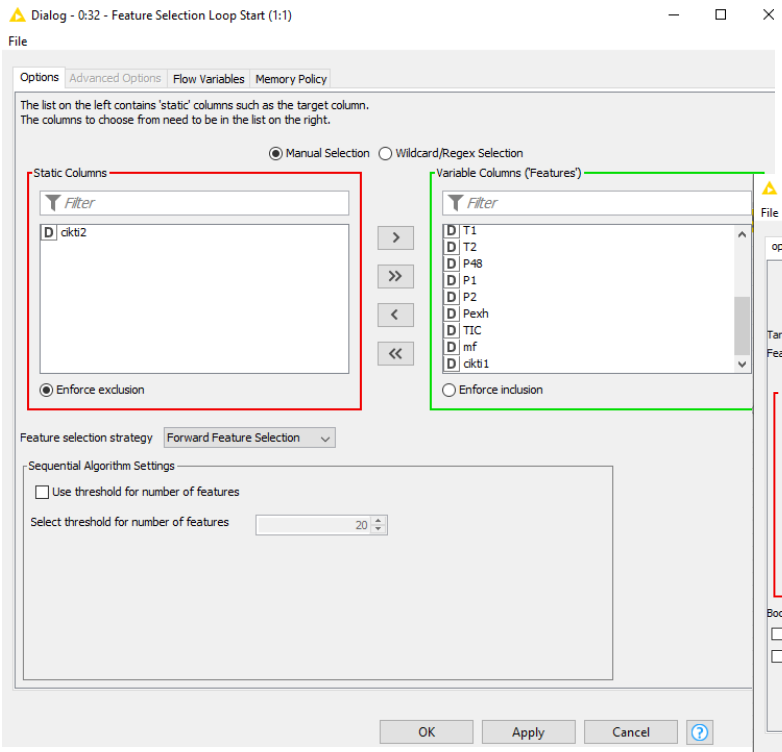
XG Boost Tree Ensemble (Regression) Makine Öğrenmesi Modeli

Feature Selection

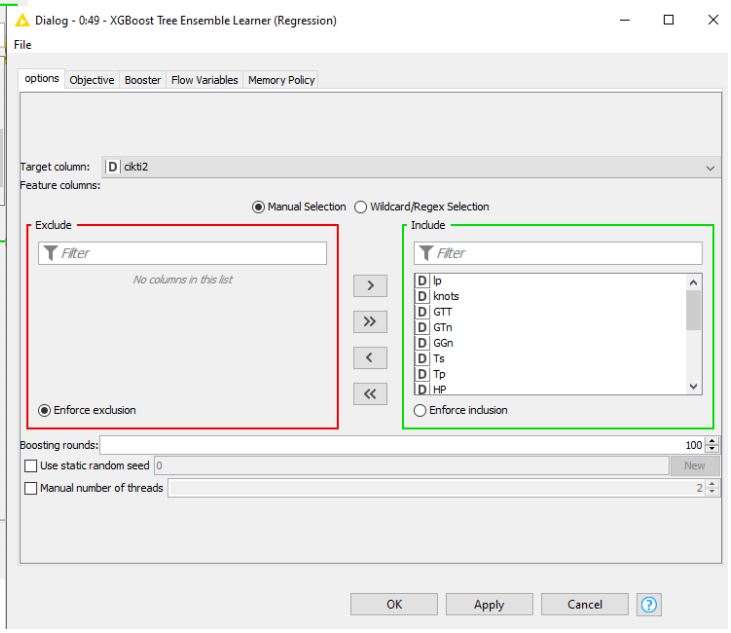
Bir önceki modelimizde sırasıyla uygulanan bütün adımlar bu makine öğrenmesi modeli için de uygulandı. Partitioning'den çıkan eğitim verileri Feature Selection Loop'a aktarıldı.

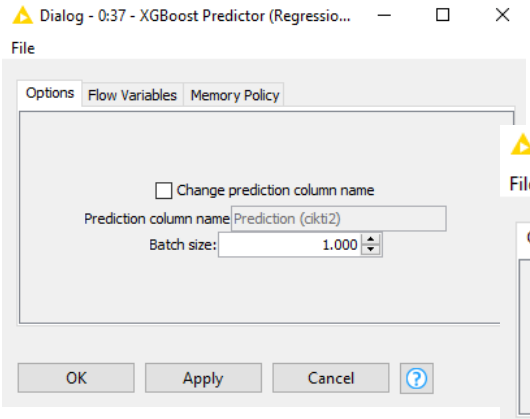


Resimde görüldüğü gibi bir akış şeması oluşturulduktan sonra yapılan ayarlar aşağıdaki gibidir.

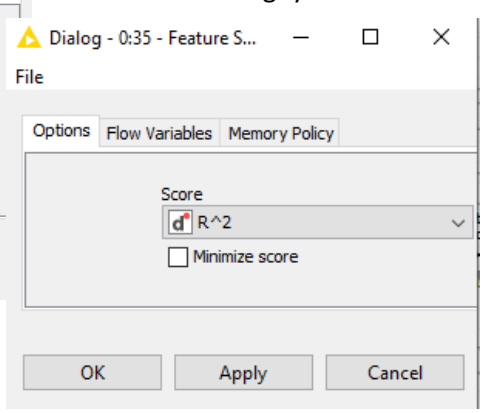


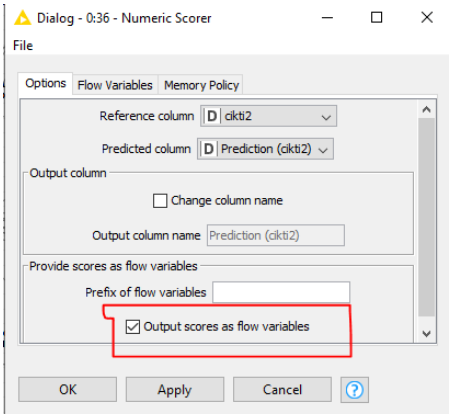
Feature Selection özelliklerinde önceki modeldeki gibi sadece çıktı değişkenimiz olan "cikti2" modelden çıkarılarak bir döngü başlatılması istendi. Modelin Learner Node'u için özelleştirme yapılmadı çıktı değişkeni kontrol edilip sonlandırıldı.

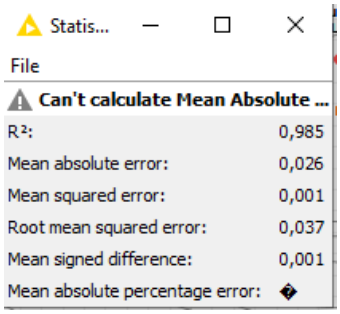




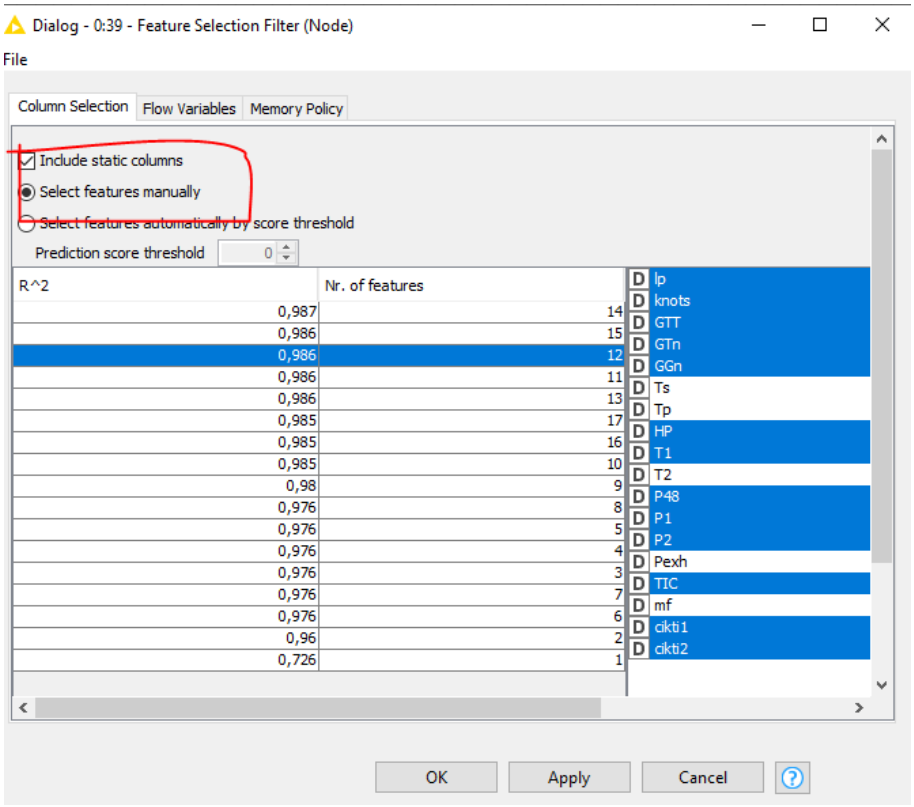
XG Boost Predictor için çıktı değişkeni kontrol edilip ayarları sonlandırıldı. Döngüyü sonlandırırken öznitelik seçiminde R^2 başarı katsayısı maximum yapacak olan değişkenlerin seçileceğini tanımladık. scorer ayarları aşagıdaki gibidir.

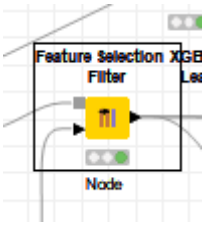






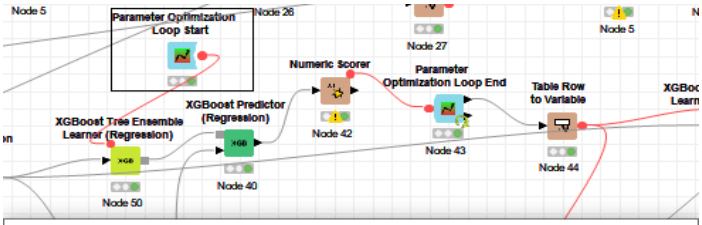
Döngü sonu scorer görünümü yandaki gibidir.





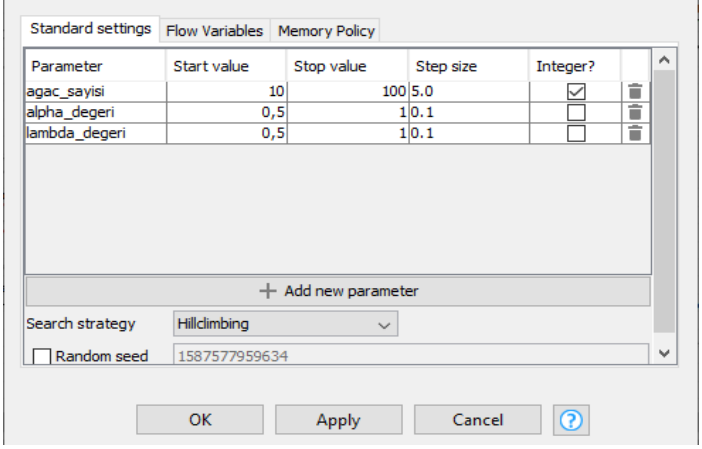
Feature Selection Filter'ın içinde döngü sonunda ortaya çıkmış öznitelikleri görüntülüyoruz. 12 özniteliğin bulunduğu tahmini R^2 Score'unun 0,98 olacağı öznitelik grubunu modelimize dahil etmeye karar veriyoruz. Bu Durumda modelimizden **"Ts, Tp, T2, Pexh, mf"** değişkenlerini çıkarıyoruz.

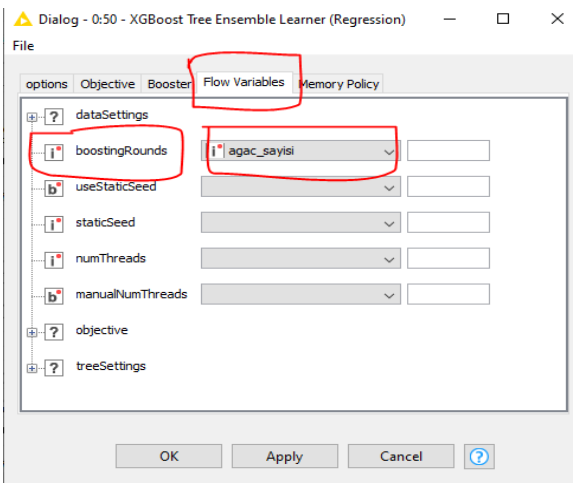
Parametre Optimizasyonu

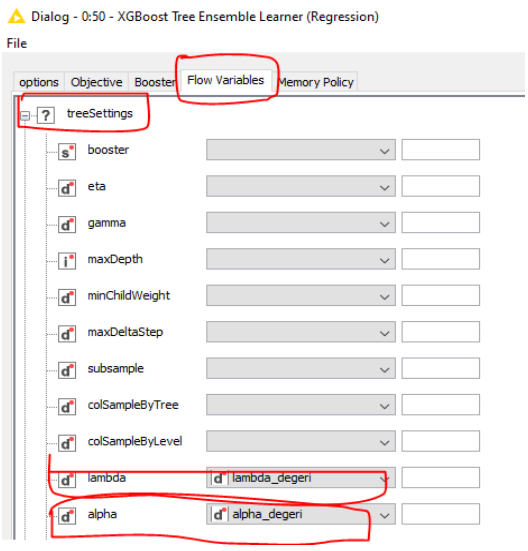


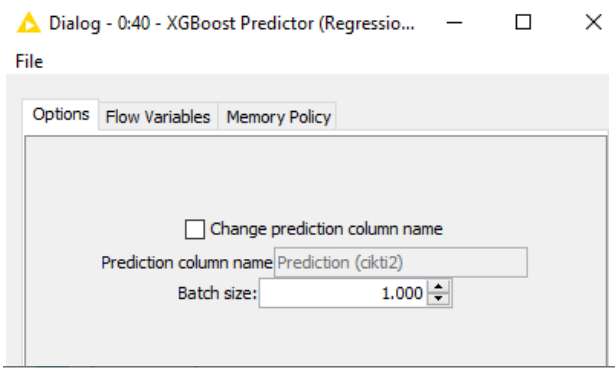
Parameter Optimization loop start ayarlarını 3 özel parametre değeri vererek yapıyoruz. İlk parametremiz XG Boost algoritmasının parametrelerinden olan **ağaç sayısını** 10 ile 100 arasında modelin deneyip en uygun ağaç sayısını bulmasını istiyoruz. İkinci parametre **alpha** değerinin 0,5 ile 1 değeri arasında 0,1 arttırarak denenmesini istedik. Son parametremiz **Lambda değerini** 0,5 ile 1 arasında 0,1 değer arttırarak denenmesini istiyoruz. Ayarlarını yaptığımız bu parametreleri modele aktarımını Parameter Loop start'dan çıkan kırmızı akış oklarıyla gerçekleştiriyoruz. XG Boost'ta bu akış oklarını açmak için sağ tuş yapıp

özelliğini aktif ediyoruz. Akış sırasında XG Boost'un parametrelerimizi bulabilmeleri için onların tanıtımını yapmamız gerekiyor. Ayarların nasıl yapıldığını açıklayan görseller aşağıdadır.



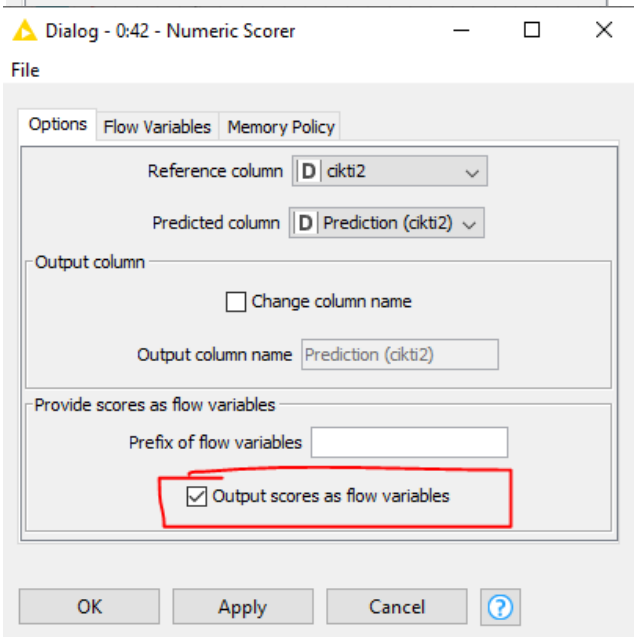






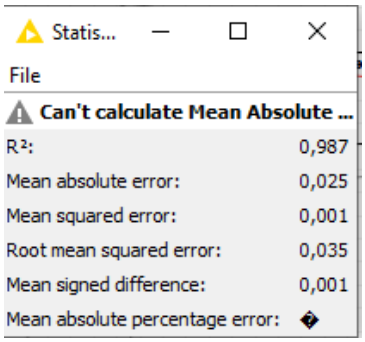
XG Boost Regression Predictor ayarlarını değiştirmedik. Çıktı değişkenin tahmin etmek istediğimiz değişken olduğundan emin olduktan sonra ayarları sonlardırdık.

Tahmin işlemi tamamlandıktan sonra scorer ayarları ve skor değerilerinin görüntüsü aşağıdakiler gibidir.

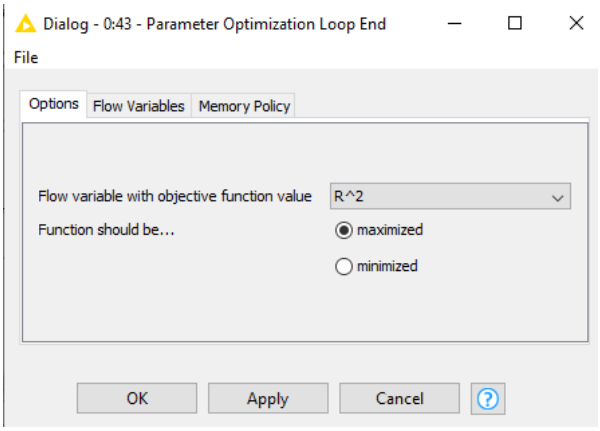


Kırmızıyla işaretlenmiş özellik parametre optimizasyon döngüsünde Scorer'dan çıkan kırmızı akış okunun parameter loop end'e veri geçişi olmasını sağlamaktadır.

Skorların Görünümü

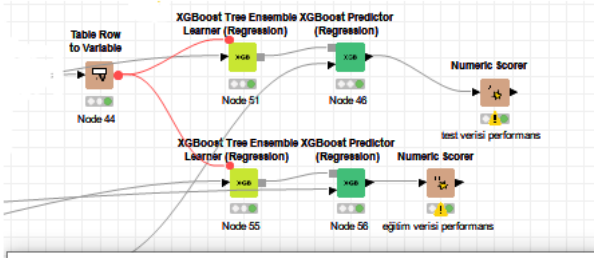


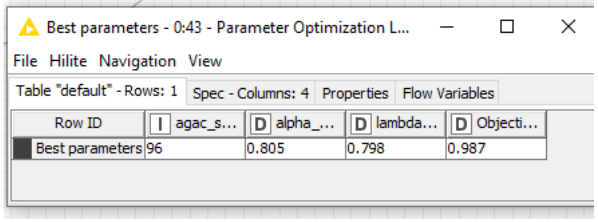
Başarı ölçümü olarak değerlendirilen R^2, **Feature Selection Loop** sonuçlarında öngörüldüğü gibi 0,98 çıkmıştır.



Parameter Optimization Loop End'i R^2 oranının en maksimum değer elde edilebilecek şekilde parametreleri bulması için ayarlıyoruz.

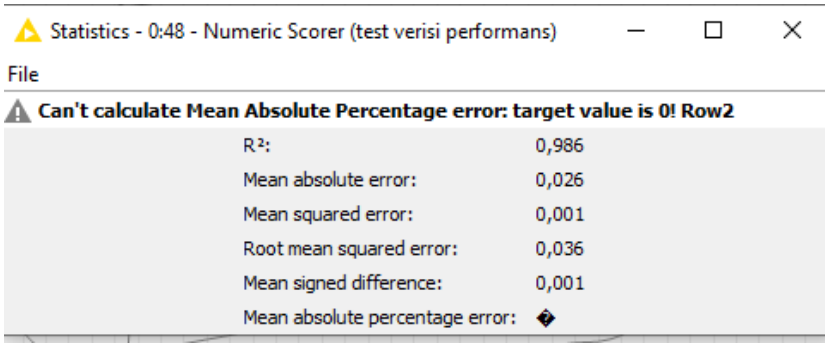
Döngü sonundaki en iyi parametreler **Table Row to Variable'ı** kullanarak işlemimizin sonuna yaklaşıyoruz. **Flow Variable** ayarlarıyla en iyi parametreleri bir sonraki aşamaya taşıyoruz. En iyi parametreleri ve önceden yaptığımız öznitelik seçimini kullanarak modelimizin test ve eğitim performansını ölçüyoruz.





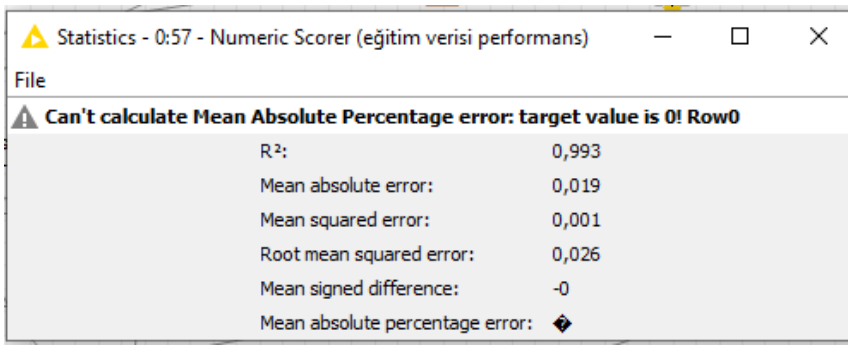
XG Boost Learner ve Predictor Node'ları üzerinde tahmin kolonumuzu kontrol etmek dışında bir işlem yapmıyoruz. Akış diyagramını yanda gördüğünüz şekilde düzenledikten sonra skor değerleri aşağıdaki gibidir.

Test performansı



Modelimizin hiç görmediği verileri tahmin etme başarısı R^2 0,98 dir. Bu modelimizin %98 başarılı olduğunu göstermektedir.

Eğitim Performansı



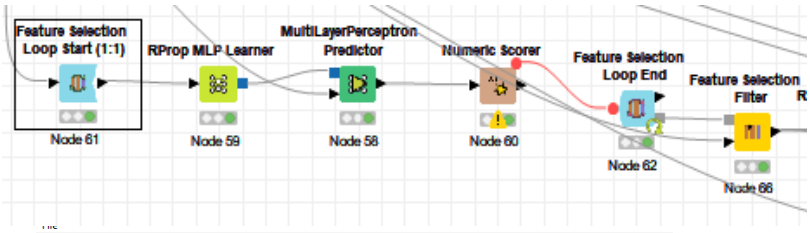
Veri setinin öğrendiği aynı veriler modele tekrar verilerek tahmin edilmesi beklenmiştir. Tahmin sonucu R^2 0,99 dur. Modelin Eğitim performans başarısının %99 olduğunu göstermektedir.

Sonuç

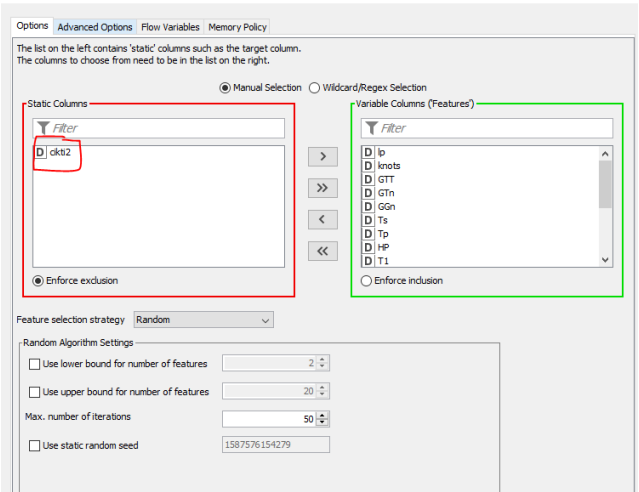
Modelde test ve eğitim performansları değerleri arasındaki farkın yok denecek kadar az olması modelin eğitimi sırasında sorun olmadığını göstermektedir. Buna rağmen değerlerin çok yüksek olması büyük olasılıkla veri setindeki her bir kolonun birbiriyle fazlasıyla ilişkili olmasından kaynaklanıyor. Korelasyon matrisinde de gözlemlemiştik bu durumu.

Yapay Sinir Ağları (Multi Layer Perceptron) Modeli

Feature Selection

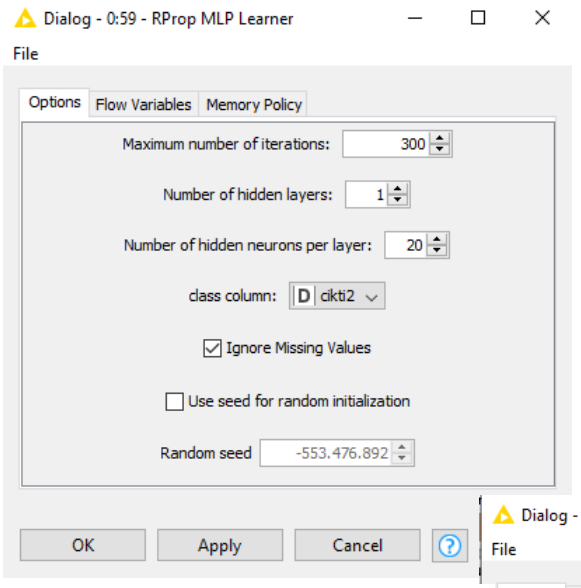


Modele başlarken **Partitioning eğitim** bölümünden çıkan akış okunu **Feature Selection'a** bağlıyoruz.



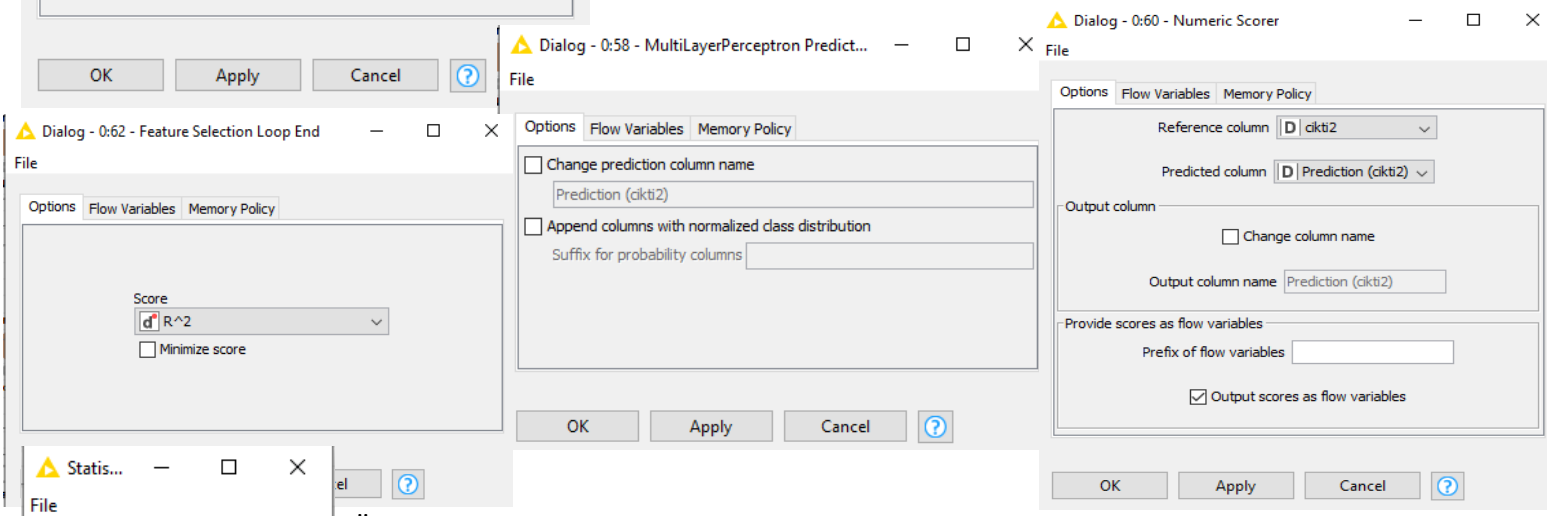
Feature Selection ayarlarında kırmızıyla yuvarlak içine alınmış tahmin kolonumuz olan "cikti 2" yi modelden çıkarmak dışında bir şey yapmadık.

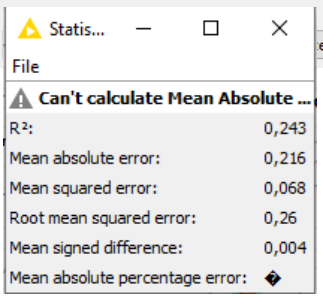
Feature Selection'dan çıkan akış okunu **Mlp Learner'a** bağlıyoruz**.**



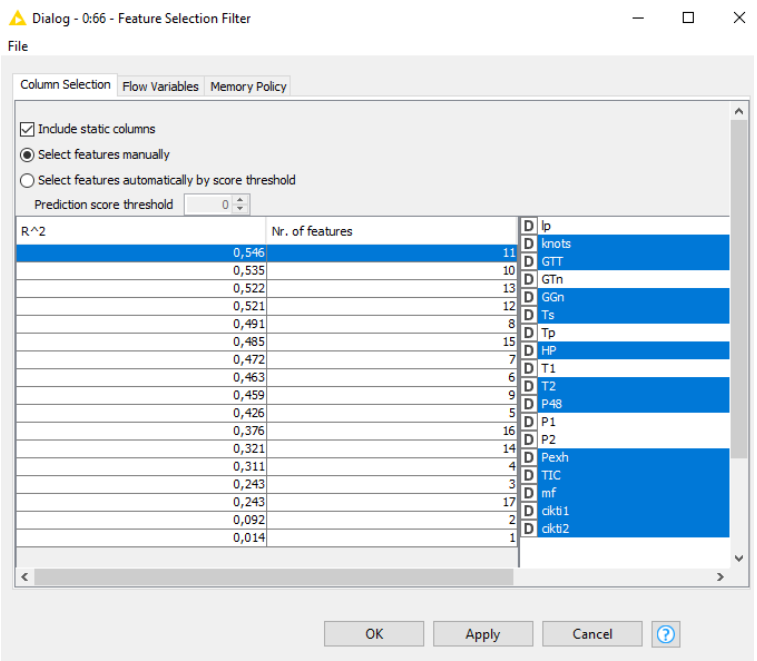
Mlp Learner ayarlarında çıktı (tahmin) olarak sınıflandırılacak kolonumuzu seçip ayarları kapatıyoruz. Mavi kutucuktan çıkan oku **MLP Prediction'a** bağlıyoruz. **Prediction** ayarlarında tahmin edilen kolonun cikti2 olduğunu belirtip ayarlarını sonlandırıyoruz. Partition'dan test vreilerini **MLP Prediction'a** aktarıyoruz. Skor ölçümü için **Numeric Scorer'ı** kullanıyoruz. Ayarlarda skorları karşılaştırılacak iki kolon olan gerçek değerler (cikti2) ve tahmin değerlerini (Prediction(cikti2)) seçerek ayarları sonlandırıyoruz. Diğer modellerde de yaptığımız gibi scorer'dan döngümüz için bilgi akışı sağlayan seçeneğimizi seçiyoruz.

Feature Selection Loop End ayarlarında R^2 skorunun seçiyoruz.



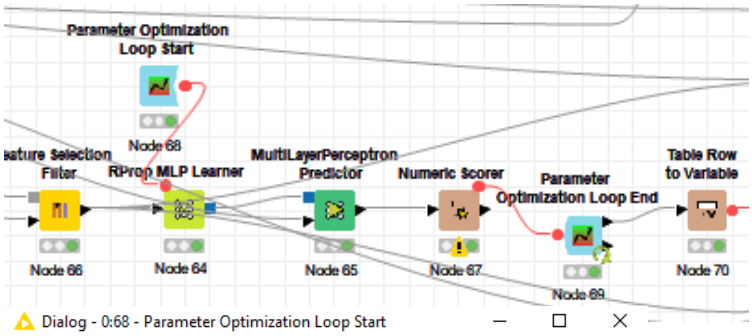


Öznitelik seçimi döngüsü bitiminde modelin scorer görüntüsü yandaki gibidir. Başarı skoru olarak ele aldığımız R^2'nin çok düşük olduğunu görmekteyiz. Seçtiğimiz öznitelikler modele dahil edildiğinde R^2 oranındaki değişikleri gözlemleyeceğiz.



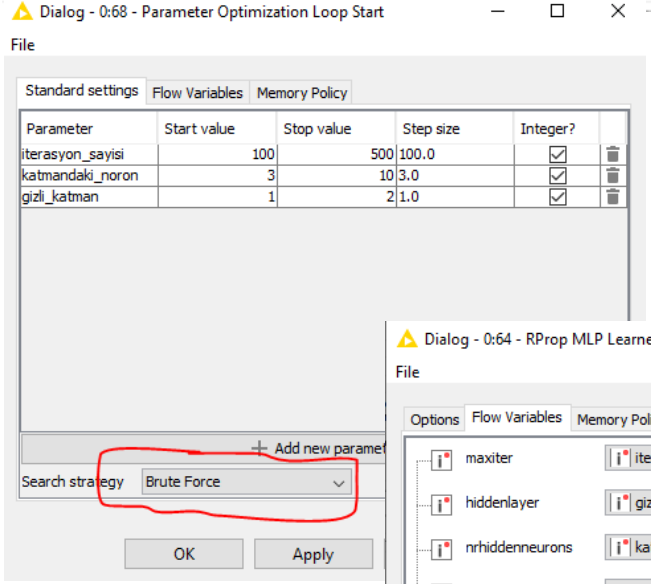
Feature Selection Filter aracını kullanarak seçilebilecek öznitelikleri aşağıda görmekteyiz. En yüksek skoru verme ihtimili olan mavi şeritle seçilmiş öznitelkleri modelimize dahil etmeye karar veriyoruz. Modelden "GTn, TP, T1, P1, P2" değişkenlerini çıkarıyoruz.

Parametre Optimizasyonu



Yapay sinir ağlarında iterasyon sayısı, gizli katman sayısı, katmandaki nöron sayısı gibi parametreler bulunur. Bizler de bu parametrelere verilmesi gereken en iyi değerleri bulup modelin başarısında olumlu etki sağlamaya çalışacağız.

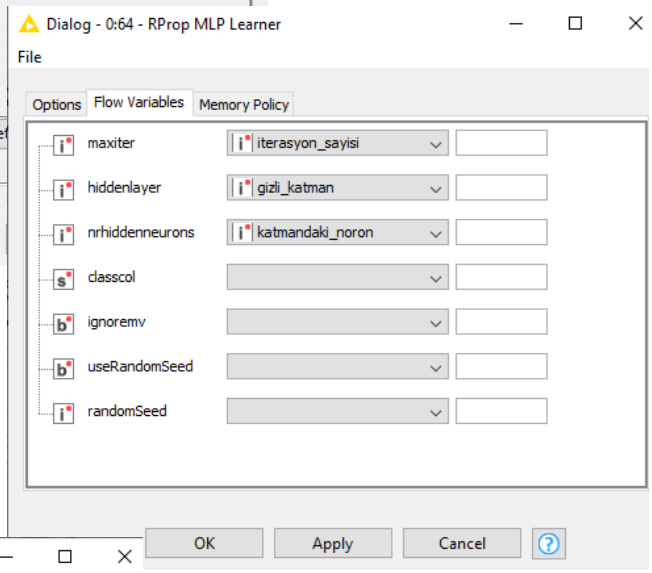
Parameter Optimization Loop Start'a kullanmak istediğimiz parametreleri temsil edecek değişkenler tanımlıyoruz.



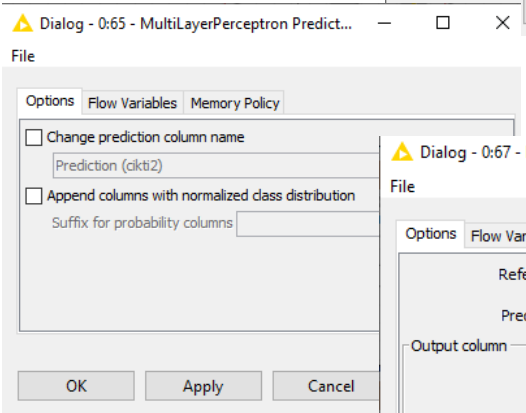
İterasyon_sayisinin 100'den başlayıp 500'e kadar 100'er artarak olasılıkların değerlendirilmesini söylüyoruz.

Katmandaki_nöron sayısının 3'den başlayarak 10'a kadar 3'er artarak denenmesini söylüyoruz..

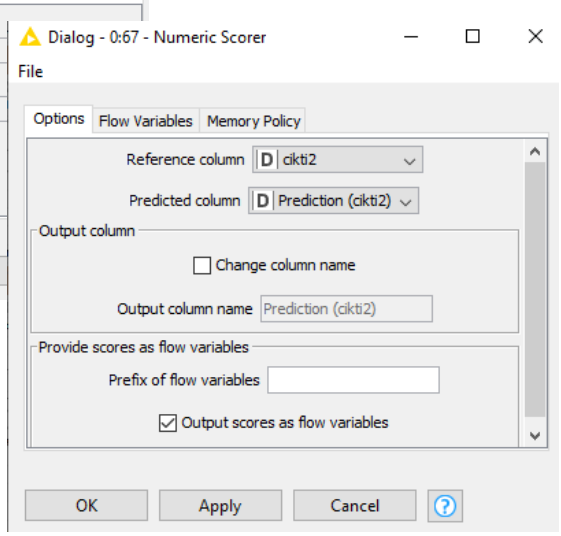
Gizli_katman sayısının 1'den başlayarak 2'ye kadar 1'er artarak denenmesini söylüyoruz.



Parametre optimizasyonu için belirlediğim değişkenlerin modelin Learner'ına bağlamak için gerekli özellikleri açtıktan sonra görseldeki gibi her bir değişkenin temsil ettiği parametreyi doğru şekilde seçiyorum.

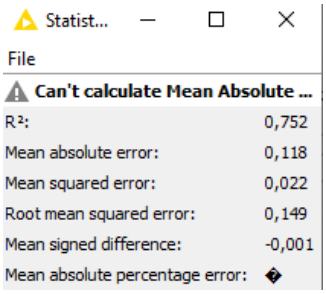


Prediction Node'unu cikti2'yi tahmin etmesi için ayarlıyor, ayarları kapatıyorum.



Scorer için de karşılaştırılacak iki kolonu seçiyor, döngüde veri akışını sağlayan özelliğini aktif edip ayarları sonlandırıyorum.

Döngü sonundaki scorer sonucunda seçilen özniteliklerin yeni modele aktarımı sonucunda modelin başarısını arttırdığını görüyoruz.



Bütün parametreler

All parameters - 0:69 - Parameter Optimization Loop End

File Hilite Navigation View

Table "default" - Rows: 30 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	iterasy...	katman...	gizli_ka...	Objecti...
Row0	100	3	1	0.104
Row1	200	3	1	0.112
Row2	300	3	1	0.191
Row3	400	3	1	0.174
Row4	500	3	1	0.281
Row5	100	6	1	0.093
Row6	200	6	1	0.104
Row7	300	6	1	0.383
Row8	400	6	1	0.253
Row9	500	6	1	0.222
Row10	100	9	1	0.068
Row11	200	9	1	0.115
Row12	300	9	1	0.203
Row13	400	9	1	0.516
Row14	500	9	1	0.217
Row15	100	3	2	0.115
Row16	200	3	2	0.237
Row17	300	3	2	0.337
Row18	400	3	2	0.423
Row19	500	3	2	0.367
Row20	100	6	2	0.135
Row21	200	6	2	0.363
Row22	300	6	2	0.311
Row23	400	6	2	0.575
Row24	500	6	2	0.564
Row25	100	9	2	0.101
Row26	200	9	2	0.513
Row27	300	9	2	0.605
Row28	400	9	2	0.688
Row29	500	9	2	0.752

En iyi parametreler

Best parameters - 0:69 - Parameter Optimization Loop End

File Hilite Navigation View

Table "default" - Rows: 1 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	iterasy...	katman...	gizli_ka...	Objecti...
Best parameters	500	9	2	0.752

En iyi parametreler Parameter Optimization loop End'den Table Row To Variable'a bağlanmış bir sonraki modele aktarılabilecek hale getirilmiştir.

Eğitim performansı

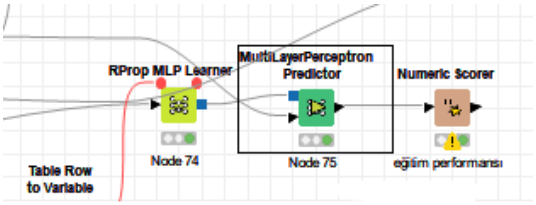
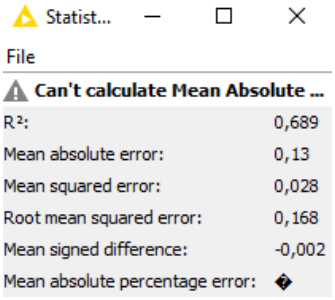


Table Row To Variable'dan MLP Learner'a parametreleri aktarıyorum. Feature Selection Filter'la seçilen öznitelikleri MLP Learner'a bağlıyorum. MLP Learner için Flow Variale tanımlamalarını yapıyorum. Eğitilmiş kısmı Predictor'a bağlıyorum. Partitioning'deki eğitim verilerini alıyor modelin tahmin etmesi için Predictor'a bağlıyorum. Predictor sonuçlarını görüntülemek için Numeric Scorer'a bağlıyorum.

Scorer sonuçları aşağıdadır.



R^2 değerinin parametre optimizasyonu yapmadan öncekine göre daha düşük olduğunu görüyoruz. Model bir önceki denemeye göre daha başarısız görünüyor.

Test Performansı

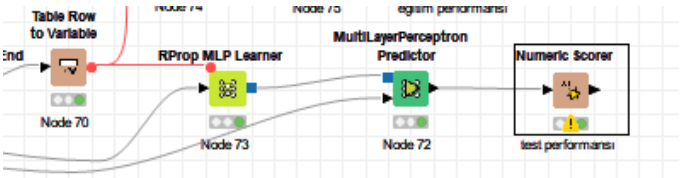
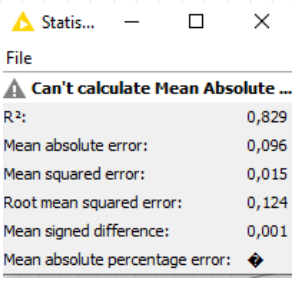


Table Row To Variable'dan MLP Learner'a parametreleri aktarıyorum. Feature Selection Filter'la seçilen öznitelikleri MLP Learner'a bağlıyorum. MLPLearner için Flow Variale tanımlamalarını yapıyorum. Eğitim kısmını Predictor'a bağlıyorum. Partitioning'deki Test verilerini Predictor'a bağlıyorum. Predictor'ı Numeric Scorer'a bağlıyor skor sonuçlarını görüntülüyorum. Test Performansı Eğitim Performansına oranla yüksek ve daha başarılı görünüyor.



Sonuç

Model daha önce gördüğü veriler üzerinden çıkarım yapması gerektiğinde %68 oranla doğru tahminler yapmaktadır. Daha önce karşılaşmadığı verilerden çıkarım yapması gerektiğinde %82 oranla doğru tahmin etmektedir.

Özet

Yukarıda ayrıntılı açıkladığım çalışmamda aynı veri seti için 3 ayrı algoritma denendi. Bu algoritmalar için ayrı ayrı öznitelik seçimi için döngüler oluşturuldu. Yine bu algoritmalar için en çeşitli parametre değerleri belirlendi ve bu değerler arasında en iyisi seçilecek şekilde bir optimizasyon döngüsü oluşturuldu. En iyi sonuçları veren öznitelikler ve parametreler yeni bir modelde her algoritma için ayrı ayrı bir araya getirilerek modelin başarısı her bir adımda daha da arttırılmaya çalışıldı. Bütün modellerin son test ve eğitim performansları karşılaştırıldığında en başarılı model sırasıyla;

1. **Polinomal Regresyon** için Test performansı: 0,99 - Eğitim Performansı: 0,99
2. **XG Boost Regresyon** için Test Performansı: 0,98 - Eğitim Performansı: 0,99
3. **Yapay Sinir Ağları (MLP)** için Test Performansı: 0,82 – Eğitim Performansı: 0,68'dir.

Modelin Tamamının Görüntüsü