

DOĞRULAMA TİPİNİN MODEL DOĞRULUK TAHMİNİNDEKİ ROLÜ

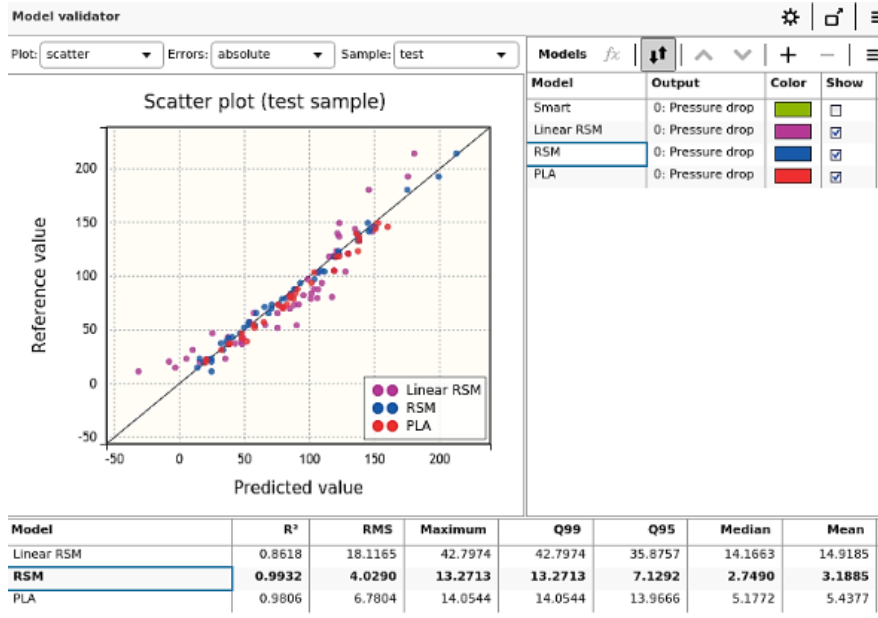
ORJİNAL METİN	HAZIRLAYAN
YULIA BOGDANOVA APPLICATION ENGINEER	UMUT CAN SALMAN MEKANİK SİMÜLASYON MÜHENDİSİ

Tarih: 02/01/2023

Simülasyon dünyasında makine öğrenmesi ve yapay zeka uygulamalarının yaygınlaşmasıyla tahmini model oluşturma, doğrulama ve hesaplama modelinin yerine kullanma eğilimi artmaktadır. pSeven tahmini model konusunda kullanıcılarına modelleme, doğrulama ve dışarı aktarma konusunda araçlar sunmakta, sürekli güncellenen kütüphaneleriyle günceli takip etmektedir. Tahmini model oluşturmak önemli olsa da doğru metrikler ile o modeli değerlendirmek, modelin hata payını ve hassasiyetini gözlemlemek oldukça önem arz etmektedir. Doğrulamanın nasıl yapılması gerektiğini saptamak için doğrulama türlerinin temellerini anlamak oldukça önem arz etmektedir. pSeven içerisinde bu konuda kullanılabilecek Model Doğrulama (Model Validator) aracı bulunmaktadır. Model kalitesini tahmin etmek ve oluşturulan modelleri karşılaştırmak için bir araç olan Model Doğrulama aracı, referans verilere göre modelin test edilmesine, çapraz doğrulama kullanılmasına, hata grafiklerini ve istatistikleri analiz ederek en doğru modeli bulunmasına olanak tanır.

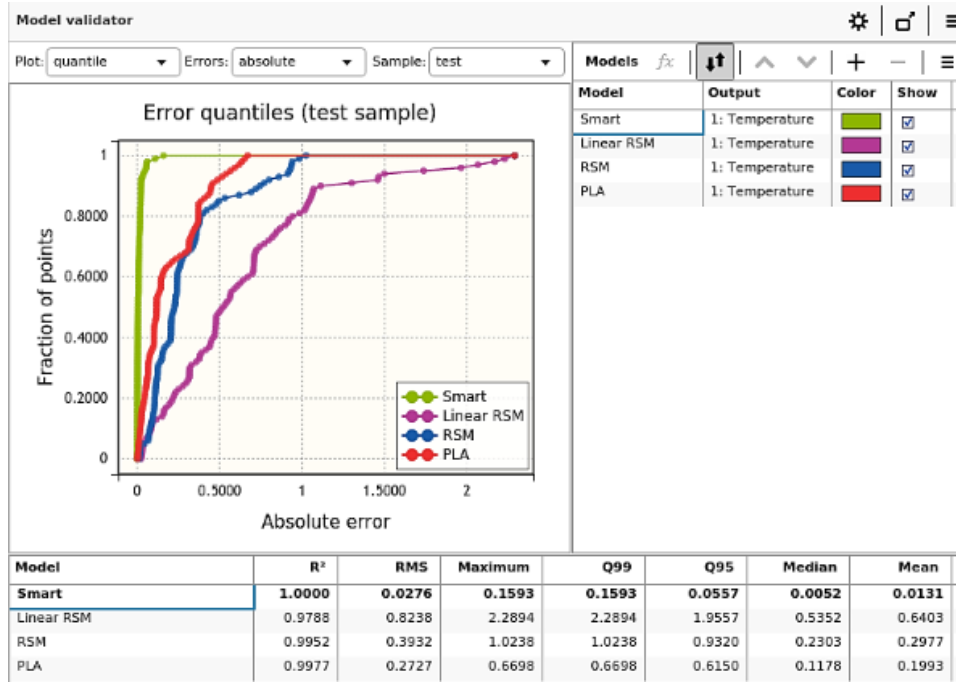
Model Doğrulama iki tür grafik göstermektedir: Bunlar:

- Dağılım grafiği, referans örnek çıktılarını doğrudan model tahminleriyle karşılaştırır (Şekil 1).



ŞEKİL 1: DAĞILIM GRAFİĞİ (SCATTER PLOT)-MODEL DOĞRULAMA

- Q-Q grafiği (varsayılan), hata dağılımını analiz etmek için kullanışlıdır (Şekil 2).



ŞEKİL 2: Q-Q GRAFİĞİ (QUANTILE PLOT) - MODEL DOĞRULAMA

Q-Q grafiğinde (Quantile Plot) her nokta, hataların yatay eksendeki değerden daha düşük olduğu örnek noktaların oranını gösterir. Örneğin, Q-Q grafiğinden alınan nokta yatay eksende 1 dikey eksende 0.80 ise, verilerin %80'ninin 1 veya 1'den az mutlak hataya sahip olduğunu gösterir. Daha dik bir eğri daha iyidir, muhtemelen grafiğin üst kısmında uzun bir yatay kısım (tail) oluşturan birkaç aykırı değerle birlikte, noktaların daha büyük bir kısmı için hata değerinin daha düşük olduğu anlamına gelir.

Doğrulama türü, model doğruluğunu tahmin etme yöntemini belirtir. Eğitim örneği (training sample), dahili doğrulama (internal validation) veya test verileri kullanılarak kalite metrikleri hesaplanabilir.

Model doğrulama için kullanılan referans verilerinin kaynağını örnek seçici (sample selector) değiştirir ve farklı şekillerde kıyaslama yapılmasını sağlar:

- Eğitim (training) seçilirse, referans veriler modelin eğitim veri setinden olacaktır.
- Test seçilirse, referans veriler test veri seti olarak beslenen veriler olacaktır.
- Dahili doğrulama (internal validation) seçeneği uygulanırsa, modelin dahili doğrulama sonuçlarından referans ve tahmin verileri okunur.

1. Eğitim (Training) Veri Seti ile Doğrulama

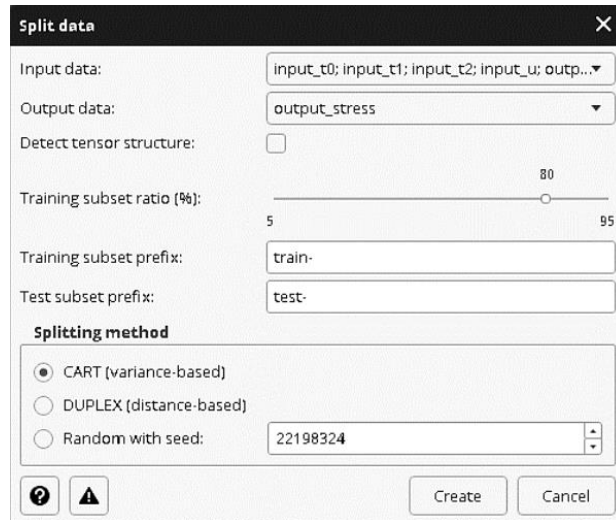
Eğitim örneği, model eğitimi için kullanılan bir veri setinden örneklenmiş bir kısımdır. Eğer eğitim veri seti referans veri olarak kullanılırsa, tahmini model hata değerlerinin küçük olması beklenir. Bu genellikle yanlış yorumlanan önemli bir gerçektir. Çoğu durumda, aşırı uyma (overfitting) nedeniyle eğitim örneğinde daha düşük hatalar ortaya çıkar ve model gösterilmeyen fakat tasarım uzayında olan noktalara uymaz veya çıktıları daha düşük doğrulukla tahmin eder. Başka bir deyişle model, eğitim verilerinin kapsamadığı, eğitim verilerine özgü davranışı öğrenmiş olur.

Bu nedenle, eğitim setinde model doğruluğunun hesaplanması zahmetsizdir, ancak elde edilen doğruluk büyük ölçüde olduğundan fazla gözükmetedir.

2. Test Veri Seti ile Doğrulama

Test veri seti, modelin oluşturulma aşamasında kullanılmayan verilerden oluşmaktadır. Bu sebeple eğitim veri setinden bağımsız olmakla birlikte muhtemelen aynı olasılık dağılımını takip eder (homojen dağılımlı bir deney tasarımı ele alınır). Bu yöntem ile Model Doğrulama aracı test girdileri için baz alınan model üzerinden tahminlerini hesaplar ve bu tahminleri test verileri ile standart hata ölçümlerini içerecek şekilde kıyaslar. Test veri seti referans olarak kullanılıyorsa, model genel olarak daha yüksek hatalar gösterir, bunun sebebi model eğitim veri setinde verilen noktalara uygundur ve bu nedenle test noktaları için tahminleri daha az doğru olabilir. Eğitim veri setine uyan model aynı zamanda test veri setine de uyuyorsa, minimum aşırı uyma (overfitting) gerçekleşmiş demektir. Bu nedenle, test örneği üzerindeki doğrulama daha güvenilir ve bilgilendiricidir. Bununla birlikte, hem test hem de eğitim veri setinin, tasarım uzayında ilgili alanı uygun şekilde temsil etmesi gerekmektedir.

Bağımsız test verileri genellikle mevcut olmayabilir, bu durumda mevcut eğitim veri setinden alt kümeler oluşturularak bir kısmını eğitim veri seti, bir kısmına da doğrulama çalışmalarında kullanmak üzere test veri seti olarak ayırmak kullanışlı olabilmektedir. pSeven içerisinde veri incele ekranında Veriyi Böl (Split Data) aracı bu işlemi gerçekleştirilebilmektedir; kullanıcı teknik ve parametre seçimi gerçekleştirebilmektedir (Şekil 3).



ŞEKİL 3: VERİ SETİNİ EĞİTİM/TEST ALT KÜMELERİNE BÖLME- DATA SPLIT

Bu seçenek, örnek verinin eğitim kümesine ve bağımsız test kümesine yüzde olarak bölünmesine izin verir. Eğitim örneğinin yüzdesi ne kadar yüksekse, model oluşturma doğruluğu o kadar iyi, ancak bu durumda eğitilmiş bir modelin doğrulanması için daha az bilgilendirici olabilir. Test veri setinin yüzdesi daha yüksek olursa oluşturulan model tahminleri kıyaslanacak test verilerine daha çok yaklaşabilir. Ancak küçük eğitim örneklemini üzerine kurulan model doğru olmayan tahminler oluşturabilir.

Varsayılan olarak, %80-20 (%80 eğitim seti – %20 test seti) bölme oranı ayarlanmıştır. Ayrıca, bir bölme yöntemi seçmek de mümkündür: CART (varyansa dayalı), DUPLEX (mesafeye dayalı) ve Rastgele.

3. Dahili Doğrulama (Internal Validation)

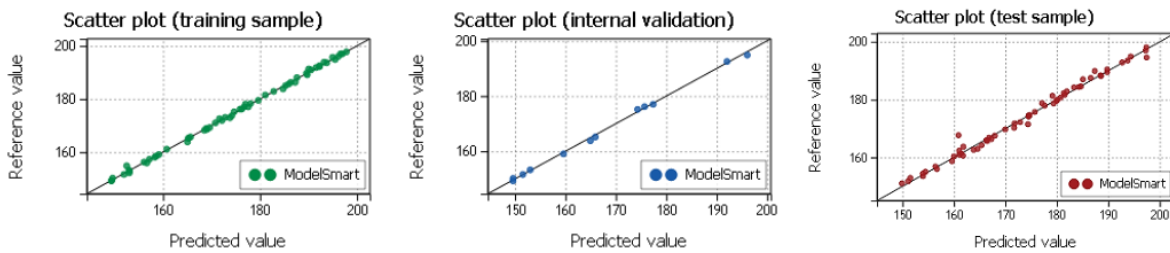
Dahili doğrulama (Internal Validation) prosedürü, tahmini model algoritmasının beklenen genel doğruluğunun bir tahminini sağlar. Bu tahmin, algoritmanın eğitim verileri üzerinde çapraz doğrulaması yapılarak elde edilir. Çapraz doğrulama, yalnızca verilen eğitim seti ile model doğruluğunun istatistiksel olarak değerlendirilmesinin iyi bilinen bir yoludur. Ancak, modelin tahmin gücünü doğrudan tahmin etmediğini vurgulamak gerekir. Çapraz doğrulamanın amacı daha çok, sonuçların tasarım alanına ölçekleyerek genişletilebileceğini ve tüm veriler üzerinde inşa edilen nihai bir modeli varsayarak, mevcut verilerin çeşitli alt kümeleri üzerinde yaklaşım algoritmasının etkinliğini değerlendirmektir.

Algoritma tüm eğitim setini alt kümelere ayırır, bir alt kümeyi test kümesi olarak ve kalanını eğitim kümesi olarak kullanır, her alt küme için tekrarlar ve ortaya çıkan hataların ortalamasını alır. Eğer test verisi yoksa dahili doğrulama daha güvenilir bir sonuç gösterir.

Çapraz doğrulama SmartSelection özelliği aktifken varsayılan olarak kullanılır.

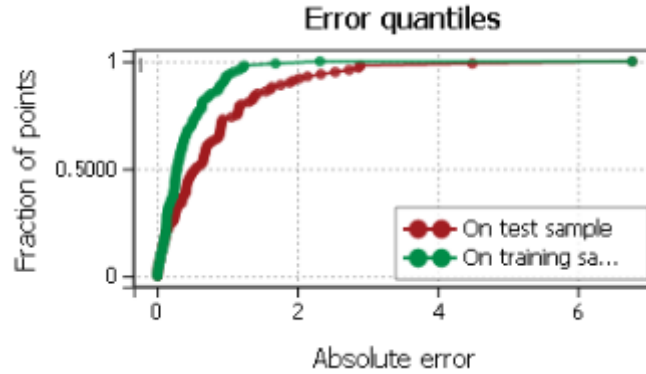
4. Kıyaslama

Bu bölümde, üç doğrulama türünün bir karşılaştırmasından bahsedilecektir. Şekil 4'te aynı çıktı için dağılım grafikleri gösterilmektedir. Referans örneğinin çıktıları ile modelin tahminleri arasındaki hatalar, noktaların dağılımı üzerinden okunabilmektedir. Noktalar ne kadar köşegen üstünde ise hata o kadar küçüktür.



ŞEKİL 4: FARKLI DOĞRULAMA TÜRLERİ İLE DAĞILIM GRAFİKLERİ

Şekil 5, doğrulama kıyaslamasında Q-Q grafiği üzerinden eğitim ve test veri setleriyle oluşturulma durumunu göstermektedir. Yukarıda bahsedildiği gibi, Q-Q grafikleri için daha dik bir eğri daha iyi olarak yorumlanabilir.



ŞEKİL 5: Q-Q GRAFİKLERİ (QUANTILES PLOTS)

Tablo 1 tahmin hatası ölçümlerini içerir. Bu ölçümlerle ilgili ayrıntılı bilgileri [pSeven yardım dokümanlarında](#) bulabilirsiniz. Eğitim veri setiyle doğrulamada hata metriklerinin değerlerinin, test veri setiyle doğrulamada elde edilen değerlerden daha küçük olduğu görülebilir.

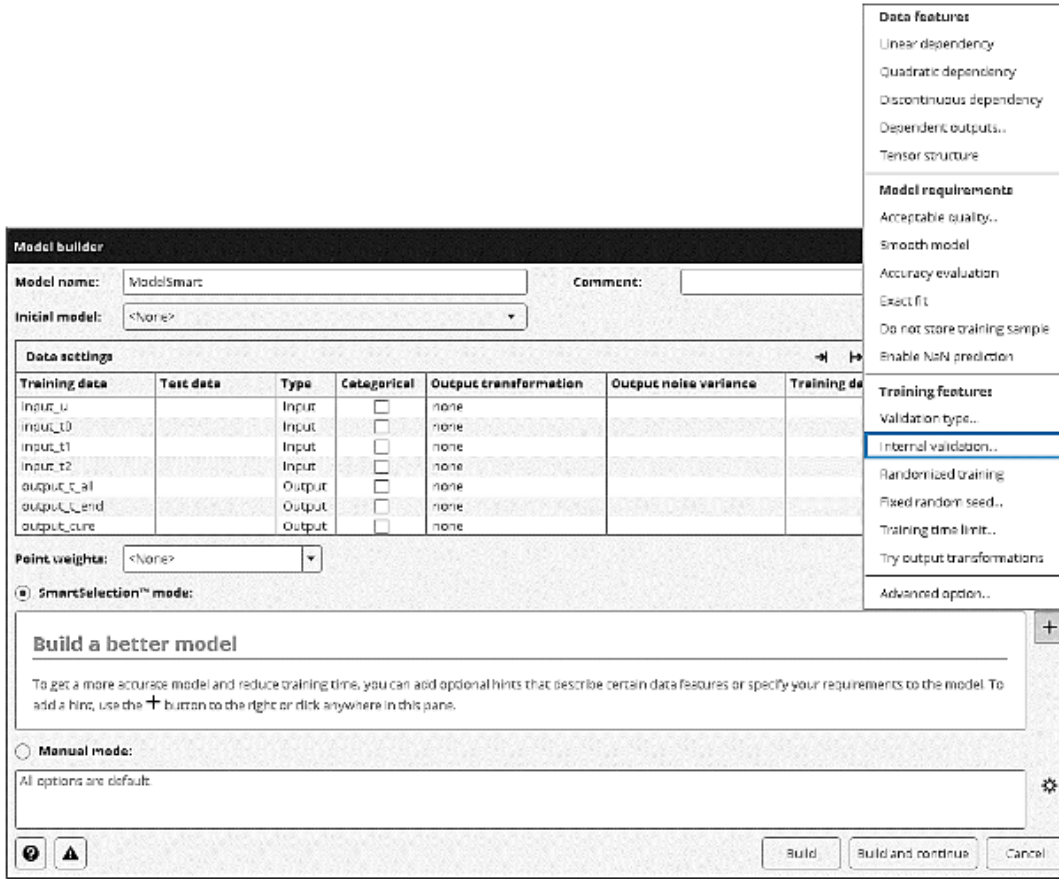
TABLO 1: FARKLI DOĞRULAMA TÜRLERİ İÇİN HATA ÖLÇÜMLERİ (MUTLAK HATALAR)

Model	R ²	RMS	Maximum	Q99	Q95	Median	Mean
On training sample	0.9985	0.5765	2.3251	2.3251	1.2151	0.2891	0.4125
Model	R ²	RMS	Maximum	Q99	Q95	Median	Mean
On internal validation	0.9980	0.6709	1.2170	1.2170	1.2170	0.5496	0.5810
Model	R ²	RMS	Maximum	Q99	Q95	Median	Mean
On test sample	0.9895	1.3447	6.7806	6.7806	2.8609	0.5722	0.8448

5. Smartselection Modunda Doğrulama Türü

SmartSelection ile model oluşturma, belirli bir problem için en doğru modeli elde etmek için otomatik olarak bir yaklaşım tekniği seçen ve çözücü yapılandırmalarını ayarlayan bir prosedürdür. Uzman olmayan kişilerin kolayca doğru vekil modeller oluşturmalarını sağlamak için tasarlanmıştır.

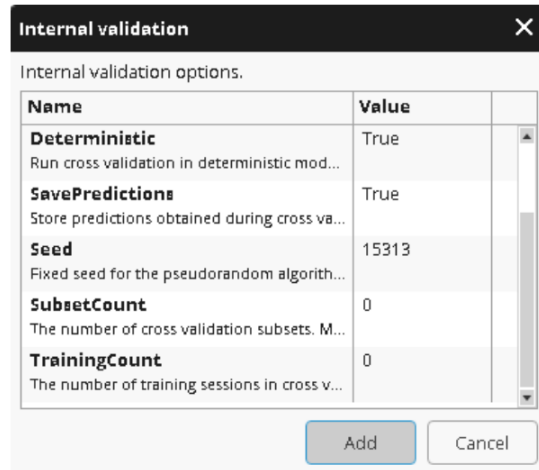
Modeli oluşturmak için gereken adımlardan biri, kalite metriklerini hesaplamak için bir örneklem hazırlamaktır. Test numunesi verilmezse, kalite metriklerini hesaplamak için çapraz doğrulama prosedürlü eğitim veri setinden örnek kullanılacaktır. Tüm dahili doğrulama seçenekleri varsayılan olarak ayarlanmıştır. Ancak, "Artı" düğmesine tıklayarak ve "Dahili doğrulama..." dizesini seçerek değiştirilebilmektedir (Şekil 6).



ŞEKİL 6: MODEL OLUŞTURUCU VE SMARTSELECTION MODU İÇİN GENİŞLETİLMİŞ İPUÇLARI BÖLMESİ

Şekil 7'de görünen yapılandırma kısmından aşağıdaki unsurlar değiştirilebilmektedir:

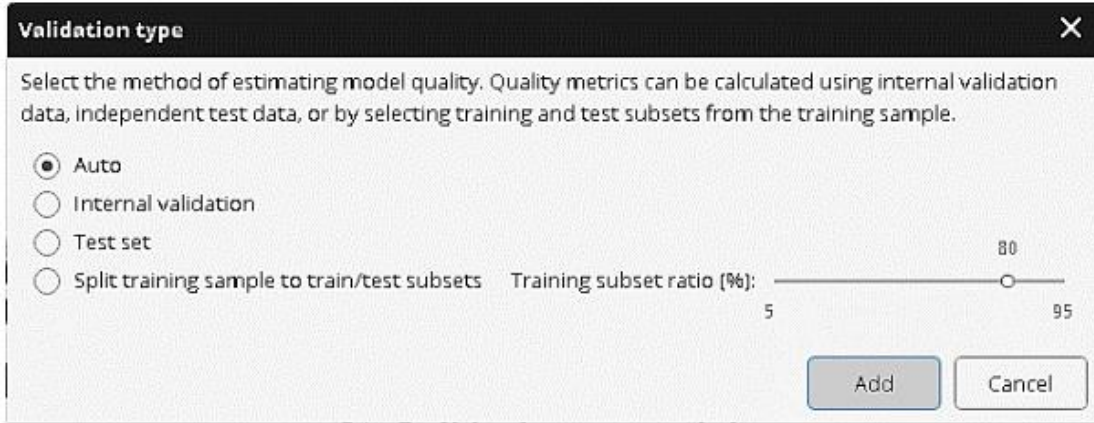
- SubsetCount – Varsayılan (0), alt küme sayısının otomatik olarak min'e (10, |S|) eşit olacağı anlamına gelir; burada |S| örneklem büyüklüğüdür.
- TrainingCount – Eğitim/doğrulama aşama sayısı. Varsayılan (0), oturum sayısının örneklem büyüklüğüne göre otomatik olarak ayarlanacağı anlamına gelir.



ŞEKİL 7: SMARTSELECTION MODUNDA DAHİLİ DOĞRULAMA AYARLARI İÇEREN YAPILANDIRMA

Çapraz doğrulama, zaman alıcı bir işlemdir, bu nedenle test verisinin kullanılması, model oluşturma süresini önemli ölçüde azaltır. Bu nedenle, test örneği mevcutsa SmartSelection bunu varsayılan olarak kullanır.

Ancak, doğrulama türü SmartSelection özelliği ile yapılan modellerde değiştirilebilir. "Dahili doğrulama..." dizisinin üzerinde bulunan "Doğrulama türü..." dizisine tıklayarak bununla alakalı ipucu sağlayabilirsiniz (Şekil 6). Şekil 8'de gösterilen yapılandırma kutucuğunda, doğrulama türünü manuel olarak ayarlayabilirsiniz; dahili doğrulama, bağımsız bir test örneği (varsa) veya belirli bir eğitim veri setinden bir test örneği elde etmek için bölme seçeneği. "Auto" tekniği seçili olduğunda, varsa test örneğini varsayılan olarak kabul eder veya yukarıda belirtildiği gibi çapraz doğrulamayı kullanır.



ŞEKİL 8: SMARTSELECTION ÖZELLİĞİ- DOĞRULAMA TÜRÜ İPUCU

6. Özet

Sonuç olarak, mümkün olduğunda modeli doğrulamak için bir test verisi örneği kullanılması önerilir. Test örneği doğrulaması, modelin eğitimde bulunmayan yeni girdi değerleri için çıktıları tahmin etme yeteneğini gösterir. Eğitim örneği doğrulaması, model doğruluğunu olduğundan fazla tahmin etme eğilimindedir. Eğitim örneğindeki düşük hatalar (daha dik hata Q-Q eğrileri), özellikle aynı model bir test örneğinde önemli ölçüde daha yüksek hatalar gösteriyorsa, gerçekte aşırı uyma (overfitting) bir işareti olabilir. Haricen test verileri mevcut değilse, dahili doğrulamaya geçilmesi önerilir, bu veriler model oluşturulurken yürütülen çapraz doğrulama testlerinden elde edilir.