**İsim :**

**Numara:**

**BLM534 Veri Madenciliği ve Teknikleri**

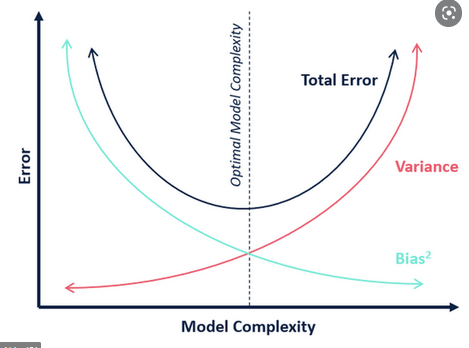
**Vize Ödevi**

1. **A**. Birliktelik kurallarında destek, güven, lift kavramlarını açıklayınız.
2. **Bread basket.csv veri dosyasında (**[**https://www.kaggle.com/mittalvasu95/the-bread-basket**](https://www.kaggle.com/mittalvasu95/the-bread-basket)**)**

Apriori, FP-Tree gibi yöntemlerden birisini kullanarak birliktelik kurallarını bulunuz. Farklı destek ve güven değerleri için analiz yaparak, sonuçları görselleştiriniz**. (25 Puan)**

1. **A.** Sınıflamada aşırı öğrenme-overfitting, eksik öğrenme-under fitting sorunlarını ve bu sorunların ne şekilde çözülebileceğini açıklayınız.

**B**. Aşağıdaki şekilde görüldüğü gibi, sınıflamada bias, variance, complexity kavramlarını ve birbiriyle olan ilişkisini açıklayınız. **(20 Puan)**



1. **A.** Dengeli-balanced ve Dengesiz-unbalanced veriseti kavramını açıklayınız. Bu durumun sınıflamaya etkilerini ve Synthetic Minority Over-sampling Technique (SMOTE) yöntemini açıklayarak, ***student performan*ce** veriseti için dengeli hale getiriniz. **(40 Puan)**
   1. <https://archive.ics.uci.edu/ml/datasets/student%2Bperformance>
   2. <https://www.kaggle.com/larsen0966/student-performance-data-set>
   3. (<https://www.researchgate.net/publication/228780408_Using_data_mining_to_predict_secondary_school_student_performance>
2. Önce dengesiz haliyle, sonra veriyi çoğaltıp dengeli hale getirdiğiniz haliyle Karar ağaçları yöntemi ve Toplulukla öğrenme yöntemleri yöntemlerinden ( Random Forest, Adaboost gibi) birini kullanarak sınıflandırma yapınız ve yaptığınız işlemleri açıklayınız.

**C.** Sınıflandırma işlemin verisetini kullanırken holdout ve cross validation için ayrı ayrı uygulamalar gerçekleştiriniz ve sonuçlarını yazınız.

**D.** 3.3.’de belirtilen yayındaki sonuçlar ile kendi sonuçlarınızı karşılaştırınız.

1. Bu ders kapsamında gerçekleştireceğiniz araştırma için seçtiğiniz veriseti, amacınız, gerekiyorsa önişleme süreci, kullanacağız yöntemi açıklayınız.

**(15 Puan)**