Nama: Kevin Gunawan  
NIM: 2201829981

1. Can you name four of the main challenges in Machine Learning?

2. What is a test set and why would you want to use it?

3. What is the purpose of a validation set?

4. What is cross-validation and why would you prefer it to a validation set?

1. – Data tidak mencukupi  
   - Kualitas data rendah  
   - Training data tidak representative terhadap case yang dikerjakan  
   - Data overfitting/underfitting
2. Test set adalah sebagian dataset yang dipisahkan dari kumpulan dataset yang dimiliki yang kemudian digunakan untuk menguji model algoritma final machine learning yang telah dibuat. Test set digunakan untuk mengetes hasil akhir algoritma agar dapat memberikan penilaian tanpa bias jika dibandingkan dengan menggunakan training set.
3. Validation set digunakan untuk mengevaluasi tanpa bias algoritma yang dibuat sekaligus melakukan tuning terhadap hyperparameters.
4. Cross- validation adalah teknik model validation untuk mengevaluasi hasil prediksi model machine learning yang dibuat terhadap dataset baru yang jarang ditemukan dalam training set guna menemukan masalah yang mungkin ada dalam model seperti overfitting atau selection bias. Cross- validation biasanya dilakukan dengan membagi validation dataset menjadi beberapa partisi yang digunakan dalam iterasi yang berbeda. Cross-validation lebih baik digunakan dibandingkan validation set normal untuk menghindari bias yang terjadi jika tedapat sebagian validation dataset yang tergabung dalam model, cross-validation juga lebih baik digunakan apabila kita tidak memiliki banyak data karena dengan cross-validation kita dapat menggunakan lebih banyak data jika dibandingkan dengan validation set.

Link Github:  
<https://github.com/nflhm/-ML_Assignment_2_Kevin_Gunawan_2201829981>