

Bitácora de experimentos

Dataset

- heart.csv

Experimento 1

Fecha y hora: 15 / julio / 2021

Especificaciones

- **Especificación 1:** Sin hiper parámetros.
- **Especificación 2:** test_size = 0.25 → De 1,025 registros, 768 fueron para entrenamiento y 257 para la evaluación.

Algoritmos

Algoritmo: Clústeres por vecinos más cercanos

Configuración:

- `KNN = KNeighborsClassifier()`
- `KNN.fit(X_train, y_train)`
- `y_pred_knn = KNN.predict(X_test)`

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 0.99
- Precisión del modelo en el subconjunto de prueba: 0.96

Algoritmo: Clasificador ingenuo de Bayes

Configuración:

- NaiveBayes = GaussianNB()
- NaiveBayes.fit(X_train, y_train)
- y_pred_nb = NaiveBayes.predict(X_test)

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 1.0
- Precisión del modelo en el subconjunto de prueba: 1.0

Algoritmo: Regresión

Configuración:

- LogisticRegression = LogisticRegression()
- LogisticRegression.fit(X_train, y_train)
- y_pred_lr = LogisticRegression.predict(X_test)

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 1.0
- Precisión del modelo en el subconjunto de prueba: 1.0

Algoritmo: Máquina de Soporte Vectorial

Configuración:

- SVC = SVC()
- SVC.fit(X_train, y_train)
- y_pred_svc = SVC.predict(X_test)

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 1.0
- Precisión del modelo en el subconjunto de prueba: 1.0

Algoritmo: Bosque Aleatorio

Configuración:

- RFC = RandomForestClassifier()
- RFC.fit(X_train, y_train)
- y_pred_rfc = RFC.predict(X_test)

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 1.0
- Precisión del modelo en el subconjunto de prueba: 1.0

Algoritmo: Red Neuronal

Configuración:

- ANN = MLPClassifier()
- ANN.fit(X_train, y_train)
- y_pred_ann = ANN.predict(X_test)

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 1.0
- Precisión del modelo en el subconjunto de prueba: 1.0

Conclusión

Los mejores modelos de clasificación son el clasificador ingenuo de Bayes, la regresión, la máquina de soporte vectorial, el bosque aleatorio y la red neuronal; sin embargo, me temo que he caído en un Overfitting.

Experimento 2

Fecha y hora: 16 / julio / 2021

Especificaciones

- **Especificación 1:** Con hiper parámetros.
- **Especificación 2:** `test_size = 0.4` → De 1,025 registros, 615 fueron para entrenamiento y 410 para la evaluación.

Algoritmos

Algoritmo: Clústeres por vecinos más cercanos

Configuración:

- (`weights = 'distance'`, `n_neighbors = 8`, `metric = 'euclidean'`)

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 1.0
- Precisión del modelo en el subconjunto de prueba: 0.98

Algoritmo: Clasificador ingenuo de Bayes

Configuración:

- NaiveBayes = GaussianNB()
- NaiveBayes.fit(X_train, y_train)
- y_pred_nb = NaiveBayes.predict(X_test)

Resultado:

- Accuracy 100%
- Sensibilidad 100%
- Especificidad 100%

Algoritmo: Regresión

Configuración:

- (penalty = 'none', C = 31622776.60168379)

Resultado:

- Precisión del modelo en el subconjunto de entrenamiento: 1.0
- Precisión del modelo en el subconjunto de prueba: 1.0

Algoritmo: Máquina de Soporte Vectorial

Configuración:

- (gamma = 0.1, C = 43939.70560760786)

Resultado:

- Accuracy 100%
- Sensibilidad 100%
- Especificidad 100%

Algoritmo: Bosque Aleatorio

Configuración:

- (n_estimators = 325, max_features = 'log2', max_depth = 10, criterion = 'entropy')

Resultado:

- Accuracy 100%
- Sensibilidad 100%
- Especificidad 100%

Algoritmo: Red Neuronal

Configuración:

- (hidden_layer_sizes = (10, 30, 10), alpha = 0.1, activation = 'relu')

Resultado:

- Accuracy 100%
- Sensibilidad 100%
- Especificidad 100%

Conclusión

Creo igualmente que he vuelto a caer en el Overfitting, pero ya se implementaron hiper parámetros... tal vez la BBDD es muy pequeña. Los mejores modelos de clasificación son el clasificador ingenuo de Bayes, la regresión, la máquina de soporte vectorial, el bosque aleatorio y la red neuronal.

Dataset Nuevo

- MiningProcess.csv

Se tomó la decisión de cambiar de proyecto y, por ende, de dataset.

Experimento 3

Fecha y hora: 05 / agosto / 2021

Especificaciones

- **Especificación 1:** Sin hiper parámetros.
- **Especificación 2:** test_size = 0.4 → De 737,453 registros, 442,471 fueron para entrenamiento y 294,982 para la evaluación.

Algoritmos

Algoritmo: Regresión lineal múltiple

Configuración con todos los datos:

- LinearRegression(normalize=True)

Resultado con todos los datos:

- Entrenamiento: MSE =2.695229066363164e+19
- Pruebas: MSE =2.683815683916006e+19

Configuración con los datos más correlacionados:

- LinearRegression(normalize=True)

Resultado con los datos más correlacionados:

- Entrenamiento: MSE =5.429073210401464e+19
- Pruebas: MSE =5.332331260275721e+19

Algoritmo: Regresión polinomial

Configuración con todos los datos:

- poly_model = LinearRegression()
- poly = PolynomialFeatures(degree=2)

Resultado con todos los datos:

- Entrenamiento: MSE =1.359648231879702e+19
- Pruebas: MSE =1.3545576697806381e+19

Configuración con los datos más correlacionados:

- poly_model = LinearRegression()
- poly = PolynomialFeatures(degree=2)

Resultado con los datos más correlacionados:

- Entrenamiento: MSE =8.37077737469533e+19
- Pruebas: MSE =8.275853203858465e+19

Algoritmo: Árbol de decisión regresión

Configuración con todos los datos:

- `adr = DecisionTreeRegressor(max_depth = 5)`

Resultado con todos los datos:

- Precisión del modelo de entrenamiento: 0.9997852290438421
- Precisión del modelo de prueba: 0.9997801911363925

Configuración con los datos más correlacionados:

- `adr = DecisionTreeRegressor(max_depth = 5)`

Resultado con los datos más correlacionados:

- Precisión del modelo de entrenamiento: 0.9997852290438421
- Precisión del modelo de prueba: 0.9997801911363925

Algoritmo: Bosque aleatorio regresión

Configuración con todos los datos:

- `bar = RandomForestRegressor(n_estimators = 300, max_depth = 8)`

Resultado con todos los datos:

- Precisión del modelo de entrenamiento: 0.9999950599159596
- Precisión del modelo de prueba: 0.9999949046646498

Configuración con los datos más correlacionados:

- `bar = RandomForestRegressor(n_estimators = 300, max_depth = 8)`

Resultado con los datos más correlacionados:

- Precisión del modelo de entrenamiento: 0.7754381533561048
- Precisión del modelo de prueba: 0.768637779618044

Conclusión

Se presentan los resultados del nuevo proyecto.