

CLASSIFICATION EXPERIMENT

Para obtener los datos de eficiencia de cada uno de los clasificadores se han obtenido las métricas de accuracy, precisión, recall y f,score.

Para que obtener datos significativos de los diferentes modelos con sus respectivos parámetros todos se han realizado mediante validación cruzada teniendo en cuenta 10 divisiones de manera estratificada de esta forma cada una de las particiones obtendrá una proporción más o menos similar de cada clase a la proporción de las mismas en el dataset completo

Los test se realizaron a los clasificadores: KNN, Logistic Regresión, MSV y Navie Bayes modificando en cada uno de ellos ciertos parámetros que se reflejan en las siguientes tablas.

KNN

KNN	vecinos	metrica	accuracy	precision	recall	f-score
1	5	euclidean	0,7937	0,65	0,4	0,669
2	5	manhattan	0,8063	0,6167	0,4	0,6739
3	3	euclidean	0,7456	0,425	0,32	0,601
4	3	manhattan	0,7699	0,5917	0,375	0,6471

Para este clasificador se modificaron el numero de vecinos a tener en cuenta y el tipo de calculo utilizado para obtener la distancia entre ellos podemos observar que en el caso de 5 vecinos la el mejor calculo de la distancia es Manhattan, pero en el caso de que se utilicen 3 vecinos apenas se aprecia la diferencia

LOGISTIC REGRESION

Logistic Reg	c	solver	accuracy	precision	recall	f-score
1	0,3	lbfgs	0,7949	0,6083	0,4	0,6679
2	0,3	liblinear	0,789	0,5917	0,4	0,6621
3	0,1	lbfgs	0,7952	0,5967	0,425	0,6711
4	0,1	liblinear	0,789	0,5917	0,4	0,6621

En el caso de regresión logística se modifican el valor de C parámetro llamado tasa de aprendizaje que indica la tasa de cambio de los pesos del modelo además de la fórmula utilizada. Para este caso apenas se aprecian cambios en cuanto a los resultados obtenidos.

SVM

SVM	kernel	random state	accuracy	precision	recall	f-score
1	linear	40	0,8187	0,6833	0,495	0,7113
2	linear	70	0,8187	0,6833	0,495	0,7113
3	sigmoid	40	0,7471	0	0	0,4275
4	sigmoid	70	0,7471	0	0	0,4275

NAIVE BAYES

accuracy	precision	recall	f-score
0,8132	0,6217	0,575	0,7273

Para el modelo de Naive Bayes no hay parámetros que se puedan modificar por lo que solo se puede obtener un valor.

CONCLUSIÓN

Con los datos ya expuestos dentro de las tablas solo queda verificar cual de los modelos y con que parámetros se ajusta mejor a el caso que queremos estudiar.

Para este caso nos interesa saber cuándo una pieza de corte se va a romper antes de que esto ocurra por ello es preferible que el clasificador nos indique la necesidad de cambiar la cuchilla antes de que sea necesario en vez de dejar pasar una cuchilla rota.

Si solo nos fijamos en la métrica del accuracy el mejor clasificador seria el SVM con el Kernel lineal y un random state de 40

Pero teniendo en cuenta solo este parámetro solo sabemos que el 82% de los casos acierta pero no podemos asegurar que los fallos se produzcan de igual forma para los dos casos para eso hemos calculado el f-score el cual realiza una simplificación entre los valores de precisión (elementos positivos entre los clasificados como positivos) y recall (elementos correctamente clasificados entre los positivos) por lo tanto atendiendo a esta métrica el mejor clasificador seria el naive bayes con un valor de F-score de 0.73.