Classificatore multiplo custom

Per creare il classificatore multiplo custom si è creata la classe *Ensemble\_classifier* per definire il tipo di oggetto. La classe contiene i metodi *\_\_init\_\_()*,  *fit()*, *base\_pred()* e *hard\_voting()*.

Il metodo *\_\_init\_\_()* viene utilizzato per creare i classificatori di base che verranno usati dall’oggetto. In quanto metodo costruttore, viene eseguito in maniera automatica quando viene definito un oggetto appartenente a questa classe. I modelli scelti sono: il Decision Tree Classifier, il Support Vector Classifier, ed il Naive Bayes Classifier. Il primo è stato scelto in quanto si tratta di un modello rapido e poco costoso, capace di gestire dati sia numerici che categorici. Inoltre, su un dataset semplice, come quello preso in analisi, ha delle prestazioni paragonabili a metodi più complessi. Lo SVC è stato scelto a causa della sua flessibilità nel gestire dataset con outlier sfruttando le variabili di slack per ottenere un *soft margin* che sia in grado di tollerarli. Infine, il Naive Bayes è poco influenzato sia dagli outlier che dagli attributi poco significativi per l’analisi, il che lo rende una buona opzione per il dataset. Con questi 3 modelli si crede che il classificatore multiplo riesca ad avere una precisione più alta dei singoli classificatori.

Il metodo *fit()* è usato per indurre i classificatori sul training set che riceve in ingresso, *base\_pred()* invece è usato per ottenere le predizioni dei singoli classificatori. *hard\_voting()*, infine, effettua la previsione finale del classificatore multiplo.

A screenshot of a computer

AI-generated content may be incorrect.Nella tabella seguente le prestazioni dei classificatori singoli e del multiplo.

La precisione ottenuta è maggiore rispetto a quella dei modelli capaci di gestire modelli più complessi, ma non raggiunge quella dell’albero decisionale. Sui dati grezzi, i risultati non giustificano l’utilizzo del classificatore multiplo. Si osserverà se il pre-processing sia capace di migliorarne le prestazioni.