Colocviu SII

Predictie note referate

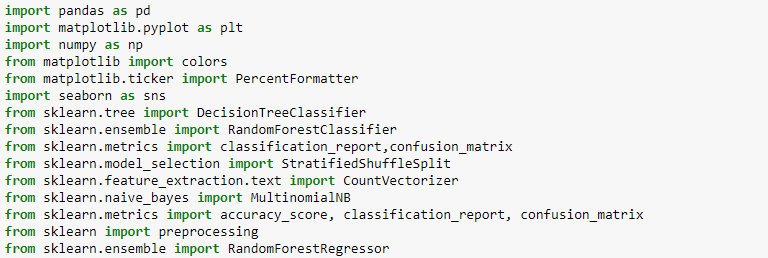
Ciolponea Eugen

Croitoru Mihai

Iordache Robert

**Descrierea implementarii**

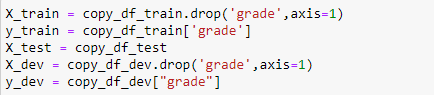
Pas 1. Importam bibliotecile necesare.



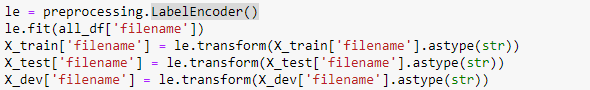
Pas 2. Importam cele 3 fisiere(train, test, dev) sub forma de dataframe-uri, folosind pd.read\_json din biblioteca pandas.

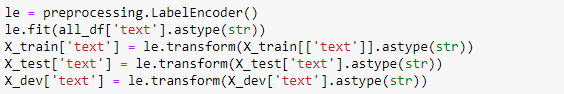


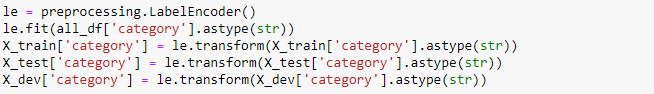
Pas 3. Pentru fiecare dataframe cream un X care contine toate datele mai putin campul grades si un Y aferent campului grades.



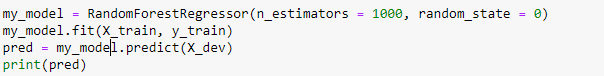
Pas 4. Aplicam functii din clasa LabelEncoder, din biblioteca sklearn.preprocessing pentru a converti date de tip string in date numerice.



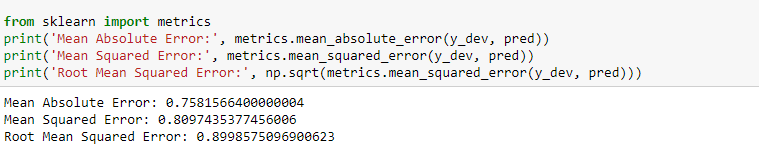




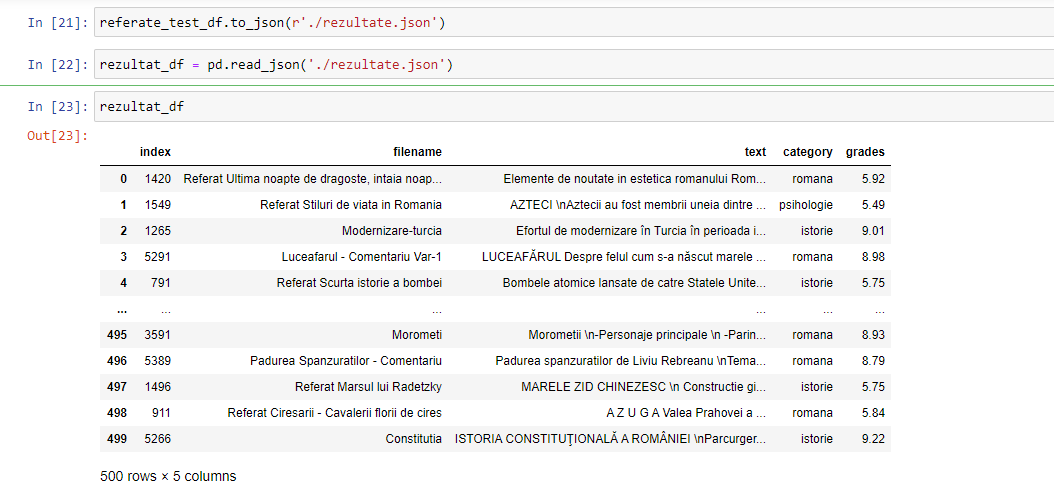
Pas 5. In cadrul acestui pas, am ales modelul folosit pentru predictia notelor referatelor din fiserul de test. Modelul a fost constituit de RandomForestRegressor din biblioteca sklearn.ensemble. S-au efectuat teste care au avut ca scop imbunatatirea acuratetii si care au constat in modificarea numarul de n\_estimator(number of trees in the forest).



Pas 6. Antrenarea a fost de tip supervizata comparandu-se rezultatele predictiilor datelor de train cu notele reale din y\_train. In urma antrenarii retelei, a fost realizata o predictie asupra datelor de dev care a fost comparata cu rezultatele reale y\_dev pentru a se face o evaluare a performantelor modelului pentru predictia notelor. Criteriul de baza pe care s-a facut evaluarea a fost Mean Square Error. In urma evaluarii s-a obtinut MSE = 0.8 , ceea ce a indicat ca modelul are performante bune.



Pas 7. A fost realizata o predictie a notelor referatelor continute de fisierul test si a fost realizat un fisier rezultate.json care contine datele din referate-test.json la care s-au adaugat notele rezultate din predictie.



**Link Github :**