

1. Raport de activare si de evaluare a performantelor

Pentru a analiza performanța s-a vizualizat un rezumat al statisticilor pentru tipurile de entități, tipurile de relații și referințele corelate. De asemenea, s-au analizat statisticile care sunt prezentate într-o matrice a erorilor. Matricea de erori ajută la compararea meta-datelor de adnotări adăugate prin intermediul unui instrument de adnotare. Metodele statistice s-au dovedit a fi o modalitate eficientă de abordare a problemelor de performanță, însă tehnicile de învățare automată funcționează deseori mai bine atunci când algoritmi sunt dotați cu indicatori care marchează ceea ce este relevant referitor la un set de date. Când discutăm despre limbajul natural, aceste indicații vin adesea sub formă de meta-date de adnotări care furnizează informații suplimentare despre text. În acest sens a fost folosit un instrument de adnotare pus la dispoziție de către platforma IBM Bluemix care furnizează următoarele metrici:

- **F1 score.** Prin folosirea acestei metrici, se ia în considerare atât precizia cât și recuperarea în vederea calculării scorului. Scorul F1 poate fi interpretat ca o medie ponderată a valorilor de precizie și recuperare, în care scorul F1 are valoarea cea mai bună la 1 și cea mai slabă la 0.
- **Precision.** Precizia specifică este fracțiunea din rezultatul instrumentului de adnotare care a fost cea mai exactă în comparație cu interpretările umane. Precizia este determinată de numărul de adnotări etichetate corect, împărțit la numărul total de adnotări adăugate de instrumentul de adnotare. Un scor de precizie de 1,0 pentru entitatea tip A înseamnă că fiecare mențiune care a fost etichetată ca tip de entitate A aparține într-adevăr clasificării respective. Scorul de precizie redusă ajută la identificarea locurilor în care instrumentul de adnotare a creat adnotări incorecte.
- **Recall.** Folosind această metrică se specifică câte mențiuni care ar fi trebuit să fie adnotate de o anumită etichetă au fost de fapt adnotate cu acea etichetă. Recuperarea este determinată de numărul de adnotări etichetate corect, împărțit la numărul de adnotări care ar fi trebuit să fie create. Un scor de recuperare de 1.0 înseamnă că fiecare mențiune care ar fi trebuit să fie etichetată ca entitate de tip A a fost etichetată corect. Un scor redus de recuperare ajută la identificarea locurilor în care instrumentul de adnotare al mașinii nu a reușit să creeze o adnotare pe care ar fi trebuit să o aibă.
- **Percentage of total annotations.** Această măsurătoare ajută la identificarea numărului de cuvinte care au fost adnotate cu un anumit tip de entitate sau tip de relație din numărul total de cuvinte, adnotate sau neanotate.
- **Percentage of documents that contain the type.** Această măsurătoare se face cu scopul de a arăta cât de multe documente conțin un tip de entitate sau un tip de relație dat. Valoarea obținută ajută pentru a vedea dacă documentele din set reprezintă suficient de bine domeniul țintă. Dacă procentul este redus pentru tipurile de entități cheie, atunci este posibil să fie nevoie să se adauge mai multe documente.

Dat fiind specificul "pilotării" în două etape, această primă evaluare nu beneficiază de volumele mari de date ("big data") care să pună în valoare capacitatea "cloud data analytics" (pentru generarea de modele comportamentale "behavioral models" individualizate – prin mecanismele inteligenței artificiale). Totodată, deși au fost activate toate funcționalitățile de bază precum și capacitățile de colectare, transmisie, stocare și procesare în cloud, nu sunt încă suficiente statistici pentru regimul de sarcină "load & stress".

