

1 Introducere

Recent, rețelele neuronale artificiale și termenul *deep learning* sunt din ce în ce mai întâlnite în învățarea automată, însă acestea nu sunt nicidecum tehnici noi, ele fiind propuse imediat după Cel de-al Doilea Război Mondial. Mai exact, prima rețea neuronală a fost construită în anul 1948 și a încercat să propună un model matematic pentru modul în care funcționează neuronii biologici. Marile piedici pentru deep learning în acea perioadă erau complexitatea computațională în procesul de antrenare și necesitatea unui nivel mare de date de antrenament pentru a obține o performanță bună. Astfel, rețelele neuronale și-au pierdut din popularitate, fiind preferate alte metode de clasificare precum SVM (Support Vector Machines) sau clasificatori liniari. Odată cu creșterea în popularitate a internet-ului, a crescut și nivelul de date distribuite public, făcând tehnicile de deep learning viabile în contextul actual, cu performanțe chiar mai bune decât metodele clasice

Domeniul *Computer Vision* a cunoscut un adevărat progres în jurul anului 2012 atunci când Alex Krizhevsky, Ilya Sutskever și Geoffrey Hinton au construit ceea ce s-a numit AlexNet. O rețea neuronală convoluțională prin care au obținut o eroare de 15.3% în cadrul competiției ImageNet LSVRC-2010 care constă în clasificarea a 1.2 milioane de imagini de rezoluție înaltă într-una din 1000 de clase. Acest rezultat a fost cu 10.8% mai bun decât precedentul, arătând că tehnicile de *deep learning* au un potențial enorm în problema recunoșterii de imagini. La data scrierii acestei lucrări, eroarea în cadrul competiției este $\approx 2.9\%$.

Prima rețea convoluțională a fost creată în anul 1998 de către Yann LeCun, numindu-se LeNet-5. Scopul ei a fost clasificarea cifrelor scrise de mână fiind inspirată de anumite descoperiri din biologie care s-au referit la faptul că, în cortexul vizual, există neuroni care răspund individual la regiuni mici dintr-un anumit stimul, creierul neprocesând o imagine ca un tot unitar.

În cadrul acestei lucrări vom studia folosința rețelelor neuronale convoluționale, a celor *obișnuite*, cât și a altor tehnici de învățare automată pentru a asigna independent unei instanțe, definită în problemă ca o mulțime de fotografii dintr-un restaurant, fiecare dintre următoarele clase:

1. bun pentru prânz (good for lunch)
2. bun pentru cină (good for dinner)

3. acceptă rezervări (takes reservations)
4. are sejur în aer liber (outdoor seating)
5. este scump (restaurant is expensive)
6. oferă alcool (has alcohol)
7. are serviciu de masă (has table service)
8. atmosfera este rustică (ambience is classy)
9. bun pentru copii (good for kids)

Această problemă a fost propusă în anul 2015 de cei de la Yelp prin platforma Kaggle. Motivația din spate a fost că răspunsurile pentru întrebările de mai sus reprezintă un factor important în sistemele lor de recomandări, însă utilizatorii nu le oferă foarte des. În acest caz, un model de învățare automată care primește ca date de intrare fotografii dintr-un restaurant și oferă o valoare din mulțimea $\{0, 1\}$ pentru fiecare clasă ar fi de folos.

Structura datelor este următoarea:

- 234842 de imagini de antrenament în format .jpg și .png.
- 237152 de imagini de test folosite pentru a determina scorul în cadrul competiției.
- ***train_photo_to_biz_ids.csv***: tabel ce asociază fiecare fotografie de antrenament la restaurantul din care provine folosind id-uri.
- ***train.csv***: tabel ce conține 2 coloane $\{business_id, labels\}$ semnificând etichetarea unui restaurant (prin id-ul său) cu o submulțime din 1, 2, ..., 9 ce reprezintă cele 9 clase ilustrate mai sus.
- ***test_photo_to_biz_ids.csv***: tabel ce asociază fiecare fotografie de test la restaurantul din care provine folosind id-uri.

În total sunt 1996 de restaurante în setul de antrenament și 10000, ce trebuie clasificate, în setul de testare.