Raportul tehnic al proiectului
"Machine learning using Cloud"
Adam Cristian A1
Pantiru Gabriel Alexandru A4

- 1. Studiu de caz
- 2. Detalii tehnice
- 3. Tehnologii folosite
- 4. Business canvas
- 5. Diagrama Arhitecturala
- 6. Diagrama Use Case

1. Studiu de caz

Proiectul nostru constă în dezvoltarea unei aplicații user-friendly pentru utilizatorii care doresc sa foloseasca servicii de clasificare date fără a fi nevoiți să aibă nevoie de cunoștințe tehnice de programare sau sa știe foarte multe informații legate de machine learning.

Ideea proiectului a pornit de la o simplă discuție cum că de ce nu am putea folosi puterea mare de calcul a cloud-ului pentru a le putea oferi utilizatorilor o aplicație care să-i ajute să-și rezolve problemele cu ajutorul învățării automate. Domenii vizate ar fi cele de cercetare științifică, dar nu pot fi excluse și alte medii de lucru unde apar probleme de clasificare care ar conduce la ușurarea rezolvării unor taskuri.

La momentul actual cred ca am găsit o nişă, de pe urma careia am putea sa ajutam si pe alții folosind serviciile și api-urile furnizate de google cloud, care duc la mărirea performanței aplicației și la reducerea riscului de incertitudine cu privire la metodele de aplicare a algoritmilor de clasificare.

Mulți consideră machine learning ca fiind un subiect foarte complicat și greu de înțeles și aplicat, iar scopul aplicației noastre este de a crea un pod de legatura, în sensul în care machine learning ar putea devine ușor de folosit pentru oricine care are un dataset deja clasificat și pe baza lui să clasifice alte elemente asemănătoare.

2. Detalii tehnice

Cum am menționat și mai sus, aplicația va fi hostata pe cloud-ul celor de la google, folosind în mare parte serviciile si api-urile bazate pe machine learning ale lor (AutoML, AutoML Tables, Cloud Vision, Cloud Natural Language, Cloud Video Intelligence). Acestea permit ca aplicația să ruleze fără vreo intarziere mai ales la crearea unui model nou pe baza dataset-ului primit de la utilizator și asigura siguranța datelor utilizatorului, baza de date fiind tot pe cloud (Firestore).

Principalele servicii furnizate de aplicatia noastra sunt clasificarea unor date pe baza unui model custom deja antrenat sau care urmează sa fie creat pe baza dataset-ului clientului și clasificarea cu module deja antrenate și furnizate direct de cloud. Pentru crearea unui model, utilizatorul furnizează un dataset de tip csv care apoi este folosit în procesul de antrenare. După ce modelul este creat folosind AutoML, utilizatorul poate sa ceara generarea unei predicții peste modelul deja creat. Pentru celălalt tip de clasificare, se vor folosi api-uri specifice precum Cloud Natural Language pentru recunoașterea entităților, Cloud Vision pentru determinare etichete imagini și Cloud Video Intelligence pentru determinare etichete videouri.

3. Tehnologii folosite

- AutoML antrenare model personalizate de învățare automata folosind-se expertiza necesara si intr-un timp cat mai scurt
- AutoML Tables antrenare model învățare automată cu date structurate
- Cloud Vision API api de detectare etichete imagine
- Cloud Natural Language API api de analizare structura și semnificația textului pentru recunoașterea entităților
- Video AI detecția etichetelor unui video
- Cloud Run platforma pentru rularea si scalerea containerelor stateless
- Cloud Functions funcții care răspund la evenimente din cloud
- Cloud Storage for Firebase depozitare obiecte pentru stocarea și servirea conținutului generat de utilizator

4. Business canvas

The Business Model Canvas

Key Partners



Principalul partener, si credem noi ca singurul, este cloud-ul celor de la google pe care este hostata aplicatia. Prin intermediul serviciilor si api-urilor furnizate de acestia permit ca aplicatia sa ruleze fara intarzieri. De asemenea, asigura siguranta datelor utilizatorului si corectitudinea raspunsurilor la cererile primite din partea clientilor.

In plus, cloud-ul mareste perfomanta aplicatiei si reduce aproape de zero riscul de nefunctionare a acesteia sau de pierdere a informatiilor acumulate. Dispare si riscul de incertitudine cu privire la metodele de aplicare a algoritmilor de clasificare.

Key Activities



Principalele activitati sunt de rezolvare a problemelor de clasificare a utilizatorilor folosind machine learning si de a le usura munca, de a le asigura corectitudenea raspunsului si cel mai important, de a le permite prin internediul unei interfete prietenoase sa poata realiza lucruri, pentru care inainte le trebuiau vaste cunostinte in domeniu.

Key Resources



Resursele de care are aplicatia nevoie pentru a ii putea oferi clientului rezultatul pe care doreste sa il afle este dataset-ul furnizat de acesta pe baza caruja este antrenat modelul. Fara model, nu se pot face predictii, deci fara dataset nu se poate construi modelul dorit de utilizator.

Value Propositions



Principalele pachete de servicii oferite de aplicatie noastra sunt urmatoarele: clasificarea unor date pe baza unui model custom deja antrenat sau care urmeaza sa fie creat pe baza unui dataset furnizat de client si clasificarea cu module deia pre-antrenate din cloudul celor de la google.

Problemele pe care le rezolva aplicatia ar fi usurinta cu care se poate folosi machine learnna fara nici un fel de cunsostinte tehnice si programare legate de ce se intampla in partea de backend. De aici rezulta o nota de convenienta si accesibilitate, dar si o asigurare a faptului ca orice problema de clasificare ar avea clientul, aceasta va fi rezolvata in cel mai rapid timp cu putinta si ca aplicatia va furniza in cele mai multe cazuri raspunsul optim, reducandu-se pentru utilizator astfel riscurile si costurile.

Customer Relationships



Customer Segments



Pe partea de comunicare cu clientii. se poate vorbi despre o asistenta tehnica dedicata si aici ne referim la postarile pe care le pot face utilizatorii pe repositoryul public din github. Ei pot face post-uri in care sa specifice ce probleme au intampinat ei in aplicatie, iar noi ne vom ocupa de ele in cel mai scurt timp posibil. De asemenea, orice client poate vedea daca problema pe care o are a mai fost discutata sau nu si daca am reusit sa o rezolvam sau daca inca lucrum la gasirea unei solutii.

Channels



Serviciile oferite de aplicatie ajung instant la clienti, imediat dupa ce stabilesc ce model vor sa foloseasca si dupa ce introduc datele pe care doresc sa le clasifice.

Folosirea serviciilor de cloud computing aduc o nota de stabilitate aplicatiei si raspuns instant la cererile utlizatoriilor, pretul fiind acelasi indiferent de timpul acestora. Deci, nu exista un serviciu cu cel mai eficient raport pret calitate, deoarece toate sunt asemanatoare

Segmentul de clienti vizati sunt utilizatorii care nu au cunostinte de programare sau tehnice despre machine learning, dar care doresc sa foloseasca algoritmul de clasificare atat in proiecte personale de invatare sau stiintifice cat si, poate, in domeniul lor de activitate.

Domenii vizate ar fi in special cele de cercetare stiintifica, dar nu pot fi excluse si alte medii de lucru unde apar probleme de clasificare care ar conduce la usurarea rezolvarii unor task-uri.

Cost Structure



Principalele costuri pe care le genereaza aplicatia sunt reprezentate de costurile de functionare a serviciilor folosite din google cloud, servicii fara de care aplicatia nu ar mai avea nici un sens. Facturarea celor de la cloud se bazeaza si pe cate request-uri se fac, deci la inceput costurile nu ar trebui sa fie foarte exagerate.

Estimam ca la inceput costurile de mentire a aplicatiei functionale in proportie de 100% vor fi mai mari decat veniturile, dar speram ca clientii vor aprecia aceasta initiativa si chiar sa se foloseasca de ea cat de mult au ei nevoie in rezolvarea propriilor task-uri.

Revenue Streams

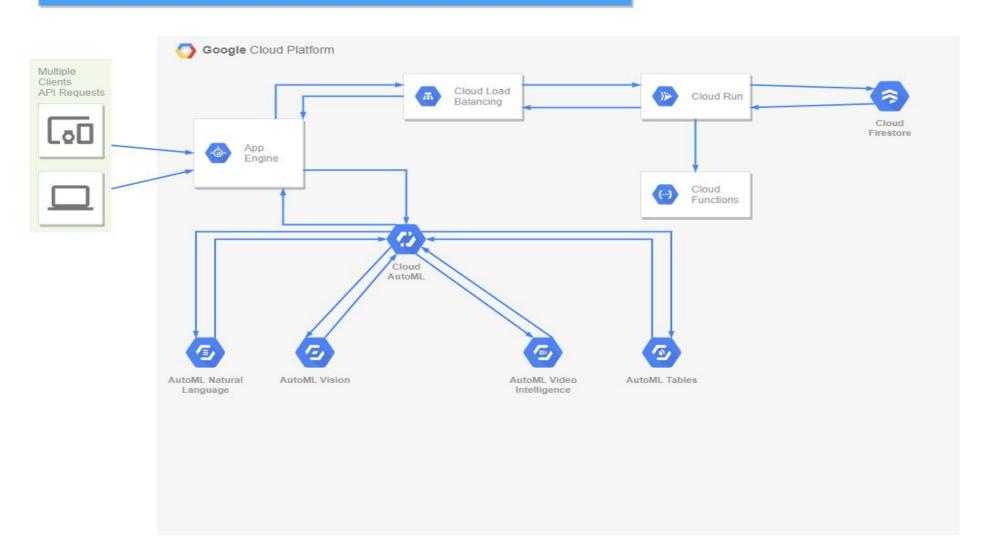


Inca de la inceput, s-a stabilit faptul ca aplicatia va o metoda de plata dinamica, mai exact fiecare utilizator plateste cate cereri face intr-o anumita perioada de timp. Din punctul nostru de vedere, e mai buna aceasta metoda flexibila decat una fixa, fiindca nu vrem sa defavorizam pe nimeni prin punerea la plata a doua persone cu aceeasi suma de bani, dintre care una foloseste aplicatia intr-un mod mai intens decat cealalta.

Vrem sa atragem constant noi clienti si ca sa reusim asta cat mai repede cu putinta, permit accesui oricui la toate functionalitatile aplicatiei, dar daca vedem ca cineva o foloseste prea intens, atunci o oprim temporar pana se rezolva aspectele financiare.

5. Diagrama arhitecturală

Architectural Diagram



6. Diagrama Use - Case

