

Predictia venitului unei persoane

Barbu Andrei-Catalin, Caragea Matei-Ioan

November 4, 2025

1 Introducere

Utilizarea pe scara larga a modelelor de machine learning (ML) in decizii critice, cum ar fi angajarea sau acordarea de credite, a ridicat ingrijorari majore cu privire la echitatea algoritmica. Modelele antrenate pe date istorice pot invata si chiar amplifica biasurile societale existente, ducand la rezultate discriminatorii bazate pe atribute demografice sensibile precum rasa sau genul. [AP]

2 Functionalitate

Proiectul propune antrenarea unui model predictiv ce are ca scop determinarea venitului unei persoane. Modelul va prezice daca venitul este mai mare sau mai mic de 50000 de dolari. Pentru acest proces vom folosi setul de date Adult Income.

Pe langa acuratetea crescuta, obiectivul proiectului este si de a urmari si asigura echitatea modelului. Vrem sa descurajam modelul din a se baza pe atribute sensibile (gen sau rasa) si a-l incuraja sa foloseasca variabilele non-discriminatorii.

3 Articole similare

Sunil T.[Tha23] a testat mai multi algoritmi de machine learning pe un set de date similar, cu 14 atribute. Comparand Logistic Regression, Random Forest Classifier, SVC si Naive Bayes, s-a observat ca Random Forest Classifier are cele mai bune rezultate, cu o acuratete de 86%. Bias-ul este abordat VS Aswathy et al. [AP], unde au observat faptul ca modelul oferea o acuratete sporita in cazul barbatilor. Diferenta dintre barbati si femei a fost micsorata utilizand tehnici de preprocessare a datelor si aplicarea metodei Exponentiated Gradient. Acesteia au reusit sa reduca diferența de acuratete de la 15.83% la 13.89%, demonstrand importanta adresarii problemei de echitate.

Lucrarea lui Chauhan [CGR] se concentreaza pe "disparitatea de echitate" (fairness disparity), ce reprezinta masura cantitativa a lipsei de paritate demografica. Raporteaza o disparitate de 0.02 in modelul de baza, care este redusa la 0.01 dupa mitigare. Prin urmare, diferența in predictiile pozitive intre genuri este acum de doar 1 punct procentual.

4 Tehnologii

4.1 Random Forest Classifier

Random Forest Classifier este o metoda de invatare automata folosita pentru performanta si rezistenta la overfitting. Aceasta tehnica functioneaza prin construirea unui numar mare de arbori de decizie in timpul etapei de antrenament si prin combinarea predictiilor acestora pentru a determina clasa finala. Prin esantionarea cu inlocuire a datelor si a selectarii aleatorii a subseturilor de caracteristicici la fiecare divizare a nodurilor, Random Forest reuseste sa imbunatasteasca semnificativ precizia si stabilitatea fata de un singur arbore de decizie. [PVG⁺¹¹]

4.2 Exponentiated gradient

Aceasta metoda de fairness adreseaza problema asigurarii egalitatii algoritmice. Bazat pe ensemble learning, EG functioneaza prin antrenarea unei sechete de clasificatori slabii, ajustand continuu ponderile asupra esantioanelor de antrenament pentru a penaliza modelul cand incalca o constrangere de echitate specifica (precum paritatea statistica sau egalitatea de sansa). Functionalitatea principala a EG este de a gasi cel mai bun compromis intre minimizarea erorii predictive si satisfacerea constrangerii bias-ului. [AP]

5 Setul de date

Setul de date ales este UCI Adult Income [Koh96], ce contine 48000 de inregistrari, fiecare descriind un individ adult prin 14 atribute demografice si economice:

1. age
2. workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
3. education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
4. education-num
5. marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
6. occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
7. relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
8. race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
9. sex: Female, Male.
10. capital-gain

11. capital-loss
12. hours-per-week
13. native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.
14. income: $> 50K, \leq 50K$

References

- [AP] VS Aswathy and Nandini Padmakumar. Fairness in predictive modeling: Addressing gender bias in income prediction through bias mitigation techniques.
- [CGR] Tejaswini Chauhan, Samarpita Ghosh, and Amrit Raj. Bias mitigation predictive modelling: A case study on income prediction.
- [Koh96] Ronny Kohavi. Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 202–207, Portland, Oregon, 1996.
- [PVG⁺11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [Tha23] Sunil Thapa. Adult income prediction using various ml algorithms. Available at SSRN 4325813, 2023.