Report Update 1

Data Collection and Use

Our company, RoadSafe Insurance, developed predictive models to predict whether a customer will file a claim and how much that claim will cost. The data collected for these models comes from previous customers who explicitly agreed, through our terms and conditions, that we could use their information for “research and development and improving current and future customer service.” We have 26 data features at our disposal, from which excluded some due to ethical concerns. We have information about our customer’s demographics (age, gender, marital status), socioeconomics (income, home value, occupation), locations household and family (number of kids), and vehicle and driving history (car age and type, travel time, etc.)

Ethical Challenge

Including sensitive attributes of customer data such as gender, marital status, single-parent status, and their location in our models poses an ethical threat to fairness and user privacy. Research by Barocas and Selbst (2016) on “big data’s disparate impact” shows how neutral features can reinforce social inequalities. They demonstrated how ML models amplified existing race and gender biases. Supporting this research, the Griggs vs Duke Power court case (1971) determined that practices with adverse effects on protected groups are discriminatory, even if unintentional.

Beyond fairness, the concept of consent complicates RoadSafe Insurance's rights to the use of customer data. Customers did sign the terms and conditions, but did they understand and were they aware that their data was going to be used to train AI models? Helen Nissenbaum would argue that data collected in one context (insurance applications) should not be repurposed in another (risk prediction) without clear explicit consent. Even though it might be legally permissible to get away with blanket consent under the constructive notice doctrine, there is still this growing concern of awareness. For instance, if AI models learn from biases, customers may experience those biases based on factors they did not realize were being used against them.

From a utilitarian perspective, including all features would yield a more accurate model, leading to lower risk, fairer premiums, and lower costs for customers. However, deontologists following the negative rights theory suggest that the mentioned features should be excluded if they risk infringing on individuals' rights not to be discriminated against based on personal characteristics. Virtue ethics agree with deontologists and further remind us that pursuing profit should not come at the expense of fairness and dignity. As mentioned, if AI systems learn and perpetuate biases that it is not able to erase, then ethical concerns arise whether consent alone is sufficient to justify the ethics behind the decisions made by AI. If consumers were aware that their consent would perpetuate biases they would probably not consent in the first place.

Balancing accuracy and protecting fairness and privacy will be a challenge. The optimal way to price insurance is indeed through data, but it might be worth it for companies to consider a small tradeoff between accuracy and ethical responsibility. In the long run, if customers feel they are being treated fairly and their information is protected, then they will remain customers for a longer period.

Standard

In our attempt to build an AI model to predict insurance claims, we always strive to uphold fairness and transparency in our processes. But historically, we discovered that demographics and gender data can be used as factors to assess how risky an insurance payout can be. This

falls in the gray zone of whether it is ethical or not because these factors can be used to create bias against certain groups in decision-making.

The ethical challenge here is that while these data points may add statistical value, they also risk reinforcing a system bias. Keeping this in mind, the direction we have chosen in our approach lessens these risks by not collecting such sensitive data and avoiding any bias altogether. The standard that we'd like to follow will eliminate any type of location-based features that might create a premeditated unfairness in our heads. GENDER, MSTATUS (marital status), PARENT1 (parental status), and Education will also be excluded as inputs to the model. Instead, we decided to only have relevant features to the predicted risk—behavioral and vehicle-specific info such as past claims frequency and vehicle type. If necessary, age and job tenure may be included if the need can be proven.

This ensures that our model upholds the industry standard in terms of accuracy of its prediction, but that it shouldn't come at the expense of bias and lack of trust.

Arguments and Objections

Our Case for Why Excluding Sensitive Features is Necessary: Excluding any kind of location-based data and demographic data is a significant step in ensuring that these AI insurance predictive models are accurate while being fair and privacy-conscious. For instance, we can look at how Amazon's AI tool discriminated against women in the recruitment cycle just because the model was trained on a feature such as the number of resumes submitted over the past 10 years. Hence, relying on relevant vehicle data instead of personal data such as marriage status and number of kids ensures that we only take into account factors that have an impact on the model's decision-making and prevent the model from developing a systemic bias.

Challenges and Counterarguments: The one objection we figured is that excluding the demographic and location factors might reduce the model's accuracy(Utilitarian approach). However, having relevant vehicle and driver-related data would be sufficient enough to discard the personal data even though they can be considered important risk factors. But in doing so we open the floodgates to an unfair predictive model. A trade-off could be to introduce an opt-in option where we provide discounts to users who are willing to provide their personal data(assuming they fully understand the implications). It is a tricky situation where it might seem like privacy-conscious individuals are being penalized but there can never be a solution- only trade-offs. Ultimately, Fair models that prioritize transparency and unbiased inputs pave the way for a more sustainable and responsible foundation for insurance pricing.

References

- Barocas, Solon, and Andrew D. Selbst. “Big Data’s Disparate Impact.” California Law Review, vol. 104, no. 3, 2016, pp. 671–732.
- Nissenbaum, Helen. “Privacy as Contextual Integrity.” Washington Law Review, vol. 79, no. 1, 2004, pp. 119–158.
- Griggs v. Duke Power Co. 401 U.S. 424. Supreme Court of the United States, 1971.
- Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". arXiv.
- Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning: Limitations and Opportunities.
- O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.