# The Ease of Inference Attacks in the Modern World

Joshua Hare
*Department of Computer Science*
*Texas A&M University*
College Station, TX, USA
jmhhare@tamu.edu

Sydney Ferris
*Department of Computer Science*
*Texas A&M University*
College Station, TX, USA
sferris@tamu.edu

***Abstract*—The proliferation of open datasets has facilitated significant advancements in data science and machine learning, yet it has also heightened concerns about data privacy and inference attacks. This study investigates privacy vulnerabilities inherent in the Adult Income Dataset from the UCI Machine Learning Repository, focusing on the ease with which sensitive attributes can be inferred from demographic data. Using machine learning models, we predict income levels (above or below $50K) and assess the dataset's susceptibility to profiling and inference attacks. The importance of each feature was calculated, and it was found that marital status and age were the most indicative of income level. Through inference attacks, we discovered that in order to have a successful inference attack there must be proper support and relevant features. However, we concluded that inference attacks are a real threat. Our findings highlight the importance of robust privacy-preserving practices to address the ethical and regulatory challenges posed by modern data sharing and analysis practices. Codebase: https://github.com/JoshHare/CSCE439ResearchPaper Video: https://youtu.be/Ia4I6y7H$_yo$**

## I. Introduction

In the digital age, the widespread availability of datasets for public and research use has raised many questions about data privacy and individual security. One such dataset is the Adult Income Dataset from UCI. This provides demographic and income-related information that can be used to train predictive models. While such datasets serve as valuable tools for machine learning applications, they can also pose privacy risks. Specifically, the data can be exploited for inference attacks, where sensitive information about individuals is deduced from seemingly innocuous demographic attributes.

This project aims to explore the privacy vulnerabilities and ethical implications of data sharing by using machine learning to predict income levels from demographic data. The study examines how demographic attributes, such as age, education, and occupation, contribute to prediction models and how they might inadvertently reveal private information about individuals. Through secondary models simulating inference attacks, we quantify privacy risks and analyze the broader implications for profiling and stereotyping.

The results of this study are particularly relevant in the context of data regulation policies such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), which advocate for digital privacy and the right to be forgotten, as well as the broader digital culture of oversharing online as a whole. By highlighting how easy it is to get private information by inferring from public information, this research advocates for responsible data practices and the integration of privacy-preserving techniques in data analytics.

## II. Related Work

The paper by Wu [3] illustrates the threat of inference attacks. This paper explores how adversaries can exploit machine learning models to infer sensitive or private information about individuals. The most effective inference methods are discussed to emphasize their effectiveness against unprotected models. It classifies the different types of inference attacks and how they are different. According to this, we will be using an attribute inference attack in our study. This paper allows us to understand the threat we are investigating and put it into a real world context.

Additionally, in a paper from IEEE [4], they dive into how data from "anonymized" data sets from governments and organizations can be cross referenced with information from social media. This can create in-depth profiles of people who were supposed to remain unknown. They similarly tried to determine what characteristics of these profiles could be used to deanonymize the data from these studies.

Finally, in a paper from Northwestern [5], they discuss how information privacy is not in danger because of individual irresponsibility or apathy, but because entities that control the data are becoming vastly more powerful than we could have predicted with less information than we thought necessary, as well as the real world implications of this reality.. Machine learning has made the personal data industry increasingly lucrative and no governing body has been able to keep regulations up to date with the current technology. The paper discusses what data is currently regulated, how individuals can control its publicity, and why that control means almost nothing with regards to their actual privacy. Their research discusses where legislation has gone right, where it has gone wrong, and what needs to be done to bring it up to date with the current technologies.

## III. Methodology

This project employs a structured approach to analyze the privacy vulnerabilities of the Adult Income Dataset and assess the risks of inference attacks:

### A. Dataset

The dataset [2] we worked with was the Adult Income Dataset from UCI. This is a multivariate dataset from the 1994 census database. The dataset has 48842 instances with 14 features each. The purpose of this dataset is to predict whether a person's income is over $50,000 a year. The features of this dataset are as follows:

- age (numerical)
- workclass
- education
- marital status
- occupation
- relationship
- race
- sex
- capital gain (numerical)
- capital loss (numerical)
- hours per week (numerical)
- native country

Once we started working with the data, we noticed that many instances were missing data fields. To fix this, we decided to drop all data that was missing features. This brought our dataset down to 6513.

### B. Random Forest Classifier

We chose to split the data with 80% for training and 20% for testing. We chose to employ a random forest classifier on all of our features. This was because it provides the feature importance scores that we needed. It also works well with categorical data.

### C. Feature Importance

Using our RFC model, the importance of each feature was calculated. The importance of the features demonstrates how impactful it is on predicting the income level. First, both the categorical and numerical features were extracted. Then, the preprocessing pipeline transforms the data by encoding it such that the classifier can use it. This is necessary because the Random Forest Classifier only takes numerical inputs.

### D. Inference Attacks

The inference attacks are simulated to predict specific target attributes from the dataset. This only works on the categorical features of the dataset. First, the data features are once again encoded, but this time one-hot encoding is used. One-hot encoding converts categorical variables into a binary matrix and ensures each category has its own feature Once again, a random forest classifier is used with a pipeline to predict the target attribute. A classification report is generated. Then, the classification report is filtered to remove classes where the support is less than 50. The purpose of this is to keep only the categorical features that have enough data to predict. Then the accuracy, weighted averages, and macro averages are recalculated to reflect these changes. After this, we will know how successful the model was at predicting other features.

## IV. Results

The base model (Table I) created with a random forest classifier had an accuracy of 83%, a weighted avg of 84%, and a macro average of 80%. The reason this is not as high as it could be is because the size of the dataset had to be severely reduced to remove data with missing rows. This was important to do because in order to accurately calculate the feature importance, we need all features to be used all of the time. It is important that our base model uses the same data that our inference attacks use in order to make appropriate comparisons.

| Label | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.94 | 0.84 | 0.88 | 4942 |
| 1 | 0.62 | 0.83 | 0.71 | 1571 |
| **Accuracy** | 0.83 | 0.83 | 0.83 | 6513 |
| **Macro Avg** | 0.78 | 0.83 | 0.8 | 6513 |
| **Weighted Avg** | 0.86 | 0.83 | 0.84 | 6513 |

TABLE I
PERFORMANCE METRICS FOR BASE RFC MODEL

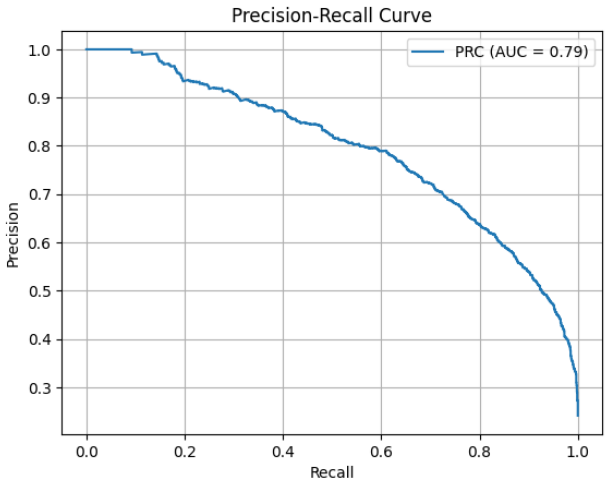


Fig. 1. Precision Recall Curve for the base model.

The three most important features of the model (Table II) were marital status, relationship, and age. A correlation between the marital status and relationship is to be expected, since this metric tracks similar information. The least important features by far were race and native country.

For the inference attacks, we only performed the analysis on categorical data. In doing the analysis, we found that many of our results were skewed by extremely imbalanced data sets. While we attempted to correct this with our pipeline, the results show that the imbalance was too big to overcome. The best example of this is with the sex feature (Table III), where there was twice as much male data than female data. Although the overall accuracy was similar to our base model, the difference in precision between male and female was staggering, with the male precision being 96% and the female precision being 66%. The support for the male set is more

| Feature | Importance |
|---|---|
| marital_status | 0.165931 |
| relationship | 0.157104 |
| age | 0.154665 |
| education_num | 0.112554 |
| capital_gain | 0.102801 |
| hours_per_week | 0.085004 |
| occupation | 0.061508 |
| education | 0.042391 |
| workclass | 0.033435 |
| capital_loss | 0.030306 |
| sex | 0.021666 |
| native_country | 0.017760 |
| race | 0.014876 |

TABLE II
FEATURE IMPORTANCE

than double that of the female set. Not only is this a large skew, it is not representative of the population sex distribution, which is much closer to a 50/50 split[1]. Even accounting for the difference in sample sizes in our pipeline, there was a huge discrepancy in our results for male and female income prediction.

| Label | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Male | 0.960091 | 0.767723 | 0.853198 | 4387 |
| Female | 0.660899 | 0.934149 | 0.774118 | 2126 |
| **Accuracy** | 0.822048 | 0.822048 | 0.822048 | 6513 |
| **Macro Avg** | 0.810495 | 0.850936 | 0.813658 | 6513 |
| **Weighted Avg** | 0.862428 | 0.822048 | 0.827385 | 6513 |

TABLE III
INFERENCE ATTACK ON SEX

The inference attack for education told a different story (Table IV). This feature had 14 categories and ended with an overall accuracy of 89.7% with a macro average of 83%. These figures are impressively high when considering that the base model was at 83% accuracy while only trying to predict between two categories. Although the data is somewhat skewed with three categories being much larger than the rest, the rest of the supports are within a similar range to each other. The precision and recall of each individual feature is still relatively high considering this imbalance. Due to the fact that predicting 14 categories is much more difficult than it is to predict two categories, this inference attack performed much better than the base model did.

However, the occupation inference attack was not nearly as impressive (Table V). The feature had 12 different categories and was largely unsuccessful in predicting them correctly with an overall accuracy of 33.3%. The support of the features was split relatively evenly between each feature, making it the most evenly distributed of all the features. We would have expected this model to do much better than it did based on this fact. No single feature reached above 60% in any metric and the lowest metrics were at 8%. We hypothesized that this is because these occupations can be held by anybody in any of the other categories, such as relationship, sex, age, and race. This makes it very difficult for the model to predict occupation based on

| Label | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Bachelors | 0.884007 | 0.955366 | 0.918302 | 1053 |
| HS-grad | 0.915736 | 0.943405 | 0.929365 | 2085 |
| 11th | 0.928571 | 0.751111 | 0.830467 | 225 |
| Masters | 0.971014 | 0.907859 | 0.938375 | 369 |
| 9th | 0.794118 | 0.704348 | 0.746544 | 115 |
| Some-college | 0.929059 | 0.917172 | 0.923077 | 1485 |
| Assoc-acdm | 0.792135 | 0.712121 | 0.75 | 198 |
| Assoc-voc | 0.800000 | 0.761905 | 0.780488 | 273 |
| 7th-8th | 0.800000 | 0.851064 | 0.824742 | 141 |
| Doctorate | 1.000000 | 0.974026 | 0.986842 | 77 |
| Prof-school | 0.944444 | 0.879310 | 0.910714 | 116 |
| 5th-6th | 0.574713 | 0.806452 | 0.671141 | 62 |
| 10th | 0.835227 | 0.803279 | 0.818942 | 183 |
| 12th | 0.779412 | 0.540816 | 0.638554 | 98 |
| **Accuracy** | 0.897531 | 0.897531 | 0.897531 | 6480 |
| **Macro Avg** | 0.853460 | 0.822017 | 0.833397 | 6480 |
| **Weighted Avg** | 0.897814 | 0.897531 | 0.896293 | 6480 |

TABLE IV
INFERENCE ATTACK ON EDUCATION

these metrics. In order to improve this model, different features would be needed that were more indicative of occupation.

| Label | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Adm-clerical | 0.363531 | 0.431129 | 0.394455 | 726 |
| Exec-managerial | 0.359467 | 0.289976 | 0.321004 | 838 |
| Handlers-cleaners | 0.168654 | 0.362637 | 0.230233 | 273 |
| Prof-specialty | 0.584527 | 0.492754 | 0.534731 | 828 |
| Other-service | 0.347475 | 0.257871 | 0.296041 | 667 |
| Sales | 0.245968 | 0.083676 | 0.124872 | 729 |
| Craft-repair | 0.288333 | 0.210719 | 0.243490 | 821 |
| Transport-moving | 0.156576 | 0.236593 | 0.188442 | 317 |
| Farming-fishing | 0.227545 | 0.393782 | 0.288425 | 193 |
| Machine-op-inspct | 0.180328 | 0.232804 | 0.203233 | 378 |
| Tech-support | 0.103550 | 0.185185 | 0.132827 | 189 |
| ? | 1.000000 | 1.000000 | 1.000000 | 389 |
| Protective-serv | 0.262712 | 0.455882 | 0.333333 | 136 |
| **Accuracy** | 0.338371 | 0.338371 | 0.338371 | 6484 |
| **Macro Avg** | 0.329897 | 0.356385 | 0.330084 | 6484 |
| **Weighted Avg** | 0.362276 | 0.338371 | 0.339458 | 6484 |

TABLE V
INFERENCE ATTACK ON OCCUPATION

The inference attacks for marital status were slightly more promising (Table VI). While never married and currently married both had high precision and statistics, the rest of the categories performed poorly in their predictions. This was likely due to the fact that their support was much lower, even 1/10th of what married and never married had. In spite of the precision being as low as 18% in some labels, the overall accuracy of the model was 83%, very close to the base model accuracy. This is to be expected as marriage status was ranked the best feature in terms of importance.

## V. CONCLUSION

The inference attack simulations yielded mixed results. With some features, such as education, we found the inference attack to be better in all metrics. In others, such as occupation, we found the inference attack to be extremely ineffective. However, the performance of these ineffective models can be explained by lack of support and lack of relevance in features. The success of the education simulation shows us that with

| Label | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Never-married | 0.85202 | 0.793509 | 0.821724 | 2126 |
| Married-civ-spouse | 0.998624 | 0.983729 | 0.99112 | 2950 |
| Divorced | 0.533333 | 0.626087 | 0.576 | 920 |
| Married-spouse-absent | 0.191489 | 0.09375 | 0.125874 | 96 |
| Separated | 0.180258 | 0.200957 | 0.190045 | 209 |
| Widowed | 0.475096 | 0.596154 | 0.528785 | 208 |
| **Accuracy** | 0.820403 | 0.820403 | 0.820403 | 6509 |
| **Macro Avg** | 0.53847 | 0.549031 | 0.538925 | 6509 |
| **Weighted Avg** | 0.830063 | 0.820403 | 0.823859 | 6509 |

TABLE VI
INFERENCE ATTACK ON MARITAL STATUS

proper support and relevant features it is possible to accurately predict private data with commonly public information. This dataset could be replicated by scraping information from Facebook, Instagram, and LinkedIn profiles. If someone were to create their own model and gather their own data, it would be possible to expose private information.

The study was limited by the dataset used. The first limitation was the lack of entries that contained every feature. This caused our total support to drop considerably from what we had initially predicted when proposing this idea. The second limitation was the continuous data. If some of the continuous data such as age were to be put into categories, we could run more inference attacks. This research can definitely be extended in the future. This could be repeated with other datasets, perhaps those with more evenly distributed support. Performing the same methodology conducted here on different data of the same type could validate or invalidate the claims made.

## REFERENCES


[1] INED, "Are there more men or more women in the world?," [Online]. Available: [Accessed: 28-Nov-2024].

[2] M. Lichman, "Adult Data Set," UCI Machine Learning Repository, 1996. [Online]. Available: https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data. [Accessed: 28-Nov-2024].

[3] F. Wu, L. Cui, S. Yao, and S. Yu, "Inference Attacks: A Taxonomy, Survey, and Promising Directions," ACM Computing Surveys, vol. 37, no. 4, August 2024. Inference Attacks: A Taxonomy, Survey, and Promising Directions

[4] J. Goldstein, "Information Privacy and the Inference Economy," Communications of the ACM, vol. 117, no. 2, 2022.

[5] J. Ferro, L. Singh, and M. Sherr, "Identifying individual vulnerability based on public data," 2013 Eleventh Annual Conference on Privacy, Security and Trust, Tarragona, Spain, 2013, pp. 119-126, doi: 10.1109/PST.2013.6596045.