PES UNIVERSITY	**PES UNIVERSITY** **(Established under Karnataka Act No. 16 of 2013)** **100 Ft. Road, BSK III Stage, Bengaluru – 560 085** **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING** **SESSION: Aug-Dec 2021**

Course Title: Ethical Algorithm Design	
Course code: UE18CS400SJ	
Semester : VII	**Team Id: 30**
SRN: PES1201800797	**Name: Harshitha Batta**
SRN: PES1201801087	**Name: Anjana V Murthy**

ASSIGNMENT REPORT

Problem Statement

There are a lot of factors that are considered for deciding which parent gets the custody of their child post a divorce. Many of these factors that are considered are sensitive. So, while building a prediction model for the same, it is important that the model doesn't defy privacy. So, a differentially private classification model for predicting the custody of a child is implemented.

Description

The dataset that has been considered for this assignment is the Mexican official government dataset for the number of divorces that took place in the city of Xalapa, Mexico. The dataset describes over 4900 divorces for the period of 15 years from 2000 to 2015 in the city of Xalapa. This dataset consists of sensitive information like Date of Birth, Date of Divorce, Place of Birth and Residence of the divorcees which is usually not found in public datasets.

The dataset was originally in spanish, which has been translated to English. It consists of the following features of the parents:

1. Date of Divorce
2. Type of divorce
3. Nationality

4. Date of Birth
5. Place of birth
6. Municipality of Birth
7. Federal Entity of birth
8. Country of Birth
9. Age
10. Municipality of the domicile where the divorced person lives
11. Federal entity of the domicile that the divorced person lives
12. Monthly Income
13. Occupation
14. Date of Marriage
15. Locality in which the marriage certificate of the divorced was drawn up
16. Municipality in which the marriage certificate of the divorced was drawn up
17. Federal entity in which the marriage certificate of the divorced was raised
18. Level of Education
19. Employment Status
20. Marriage_duration
21. Number of Children
22. Custody

A few more synthetic records have been added to create a dataset of 7924 rows. As the model is going to be built to predict which Parent will get the custody of the child, only the most important features pertaining to the parents have been selected which are Place of Residence,Place of Birth, Occupation, Level of Education, Monthly Income and Age of the parents.

As seen above there is a lot of sensitive information provided in the dataset. Attributes like age, place of residence, place of birth and a person's salary are sensitive information and should be shielded from outsiders. If used in the training dataset for a classification model, a malicious attacker can inspect the model and get to know private information about people which can lead to a compromise on safety and security of the people sharing their data. So, to avoid this, differential privacy is required.

In the model that has been built, Differential Privacy has been implemented using the private machine learning technique PATE-*Private Aggregation of Teacher Ensembles.*

In PATE, the private dataset is made into disjoint partitions and these partitions are trained on separate classifiers, called Teachers. Another classifier is built called the Student, which will be trained using data without labels. The teachers predict labels and the aggregate of those labels are considered with noise added, as the labels for the Student dataset. Finally, the student dataset is trained with the labels predicted by the Teachers. Using this method assures Differential Privacy.

In the model built,

- Number of teachers that have been considered are 6 classifiers, each trained using SVM with the kernel "poly"
- Noise that is added to the aggregate of the labels, is "exponential noise".
- The student dataset is trained using the "linear" kernel of SVM.

Output Screenshots

- Before Adding Differential Privacy,
  Time taken for training the SVM Model,

```
#Without Differential Privacy
%%time
from sklearn import svm
clf = svm.SVC(kernel='linear')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

```
CPU times: user 7min 18s, sys: 341 ms, total: 7min 19s
Wall time: 7min 17s
```

Accuracy of the SVM Model,

```
from sklearn import metrics
from sklearn.metrics import precision_recall_fscore_support
print(precision_recall_fscore_support(y_test, y_pred, average='macro'))
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

```
(0.3810949955177094, 0.37461788655265194, 0.34884312229849995, None)
Accuracy: 0.47124824684431976
```

Confusion Matrix of the SVM Model,

```
print("Confusion Matrix\n",metrics.confusion_matrix(y_test, y_pred))
```

```
Confusion Matrix
 [[ 33 119  29]
 [ 28 280  33]
 [ 34 134  23]]
```

- After adding Differential Privacy,
  Time taken for training the SVM Model,

```
%%time
from sklearn import svm
clf = svm.SVC(kernel='linear')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

```
CPU times: user 47.6 s, sys: 44.9 ms, total: 47.6 s
Wall time: 47.4 s
```

Accuracy of the SVM Model,

```
from sklearn import metrics
print(precision_recall_fscore_support(y_test, y_pred, average='macro'))
print("Accuracy:",metrics.accuracy_score(sy_test, y_pred))
```

```
(0.6904970329775861, 0.6486965987805116, 0.6645256100720559, None)
Accuracy: 0.4838709677419355
```

Confusion Matrix of the SVM Model,

```
print("Confusion Matrix\n",metrics.confusion_matrix(sy_test, y_pred))
```

```
Confusion Matrix
 [[ 10 107  64]
 [ 10 276  55]
 [  8 124  59]]
```

Table of Results Obtained,

	Accuracy	Precision	Recall	F-Score	Time
With Privacy	0.471	0.381	0.371	0.348	7 min 19s
Without Privacy	0.483	0.690	0.648	0.664	47.4 s

Interpretation of efficiency

The model has been built using the PATE framework, the intuition behind the framework is that a single person's information does not affect the outcome of the model's learning and that person's features are not memorized and the privacy is maintained. If n classifiers are trained on different disjoint sets of data, and all the classifiers agree on what the predicted label should be on a new input given, then the predicted label does not give information of a single training example.

Learning Outcome

- Choosing a dataset that contains sensitive and private information and understanding the importance of applying differential privacy to protect the information.
- Using anonymization on the dataset is not a completely foolproof method of ensuring privacy as it cannot prevent linkage attacks.
- Various methods of applying noise to the dataset like laplace noise and gaussian noise to ensure privacy.
- Intuition behind PATE method of differential privacy and its implementation
- Efficient methods to evaluate performance of models before and after applying differential privacy.

Name and Signature of the Faculty