Fairness in Intrusion Detection

Hemendra Jampala

Mukesh Muvva

GUIDANCE

PROF. MOHAMMED NASSER

PROF. MOHAMMED AMINUL ISLAM

Data Set

NSL-KDD	CICIDS2017	UNSW-NB15
KDD Cup 99	CTU13	1999 DARPA.

Dataset

- As we figured out that all the intrusion related data sets were kind of
  phony and fake. They were generated in a lab.
  • All the data sets which we were goofing around Intrusion Detection had
  nothing to do with fairness.
  • To deal with fairness of any cyberattack detection we need sensitive
  data like Age, Gender, Country, race etc..,
  • We have changed our approach in finding data sets.

AWS Honey Pot Data Set

- Lure the attacker by setting up a fake real time system
  - Honey Pot deflects your threats
  - This data set is based on AWS honey pot.

Do we have fairness data here?
Yes, But the only attribute we have here is Country, Location and IP Address

The AWS honeypot won't be enough for the task we wanted to do
- It will be mainly useful Data Exploration
- The Data Set is only of one class
- Malware Class

Caveats that apply to this dataset

This dataset and the types of worm and denial-of-service attack traffic contained therein are representative only of some spoofed source denial-of-service attacks. Many denial-of-service attackers do not spoof source IP addresses when they attack their victim, in which case backscatter would not appear on a telescope. Attackers can also spoof in a non-random fashion, which will incur an uneven distribution of backscatter across the IPv4 address space, and may cause backscatter traffic to miss any telescope lenses. Note that the telescope does not send any packets in response, which also limits insight into the traffic it sees.

Data Access Policy

These data must be analyzed on CAIDA machines, and cannot be downloaded!

Academic researchers and US government agencies can request access through CAIDA by filling out and submitting the online form. It usually takes about five to ten business days to process your request. We carefully review each application and the decision to grant the data access is based on the merits of your proposed data use.

These data also may be available for corporate entities who participate in CAIDA's membership program. Information on membership levels, services, and rates can be requested by emailing sponsorship@caida.org.

Once users are approved for access to this dataset, they will receive an account on the CAIDA machine that provides direct access to the Telescope data they requested. Accounts are valid for a nominal twelve months in which the research is expected to be completed. CAIDA strictly enforces a "take software to the data" policy for this dataset: all analysis must be performed on CAIDA computers; download of raw data is not allowed. CAIDA provides several basic tools to work with the dataset, including CoralReef and Corsaro. Researchers can also upload their own analysis software.

CAIDA-Real time network telescope data.

More Datasets

Wildcard 400 of the 2019 Trendmicro CTF

- quirks: Timestamp,src,dst,port,bytes

RoEduNet-SIMARGL2021 Network Intrusion Detection Dataset

- Quirks: Only Source

Final Conclusion on Dataset

- We need a dataset
  - Two Classes – Malware and not a Malware
  - Physical Features – Country, City, Age, Gender
  - Network Data – Features related to Malware

IoT-23 Data Set

- Captured the data in stratosphere laboratory
  - Funded by Avast Software Prague
- How to get this

  You can download them from their website

  A reference should be mentioned

- Idyosyncranacies of this dataset

  Physical features
  - Type of malware classification
  - Humongous data set

  Challenges

  - Very big data set
  - Takes lot of time to process
  - Sometimes manually have to copy some values
  - Because network data comes in PCAP file.

Data Preprocessing

- Loaded all the data into data frames
- Checked whether there are multiple countries or not
- Trying to figure out country names from IP Address using API
- Once having the country and region our data preprocessing task would be done.

Challenges

- Humongous data set
- Google collab file crashed multiple times while trying to load data
- Couldn’t find free API to convert IP address to Country names

POC’s – Data Related

Trying to save data into google drive

- Copying the data from collab to drive
- Usage of HDF5 files, pickle files as backup, so that I am saving the current state of my execution.

Model

- Implemented a Fairness model Fair Learn on our code
  Is It useful? Partially somewhat

- Continent and Country has nearly 45% influence in predicting the label
  It’s completely okay to have influence of features like missed bytes, Service
  Miscellaneous factors like Timestamp does also have influence on the model.

- Challenges
  Some important libraries were missing in the code,
  So had to work around the library

Neural Network

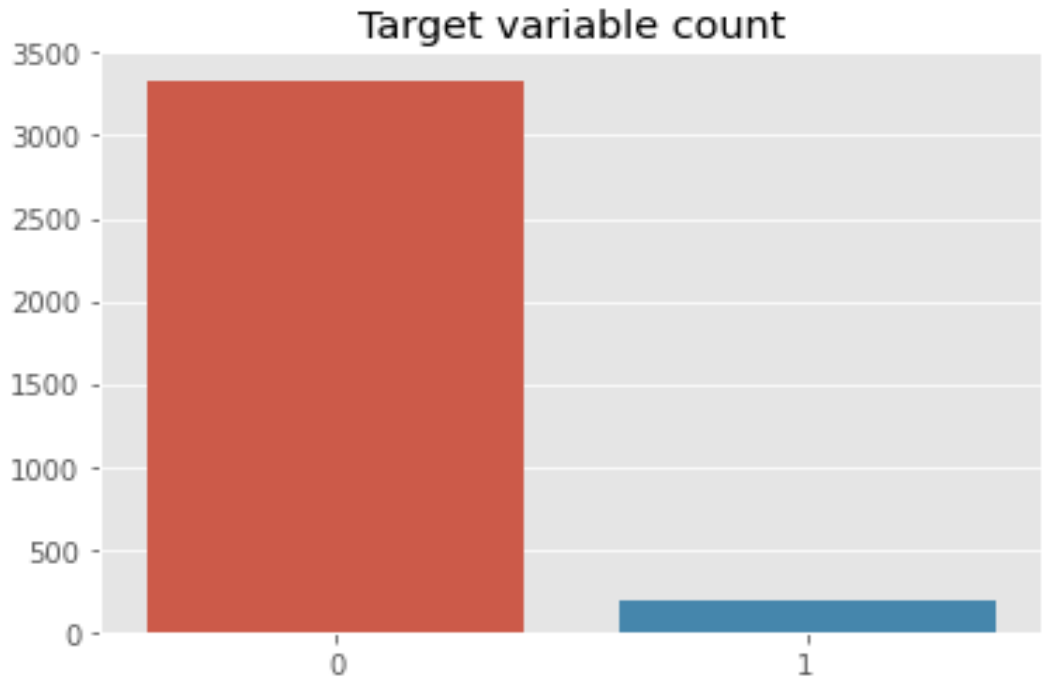
- A basic Neural Network was built on the model
- Model(
  (layer1): Linear(in_features=15, out_features=50, bias=True)
  (layer2): Linear(in_features=50, out_features=50, bias=True)
  (layer3): Linear(in_features=50, out_features=3, bias=True))

- RESULTS
  For ML (KNN) - Accuracy :95%
  For Neural Network – Accuracy :94%

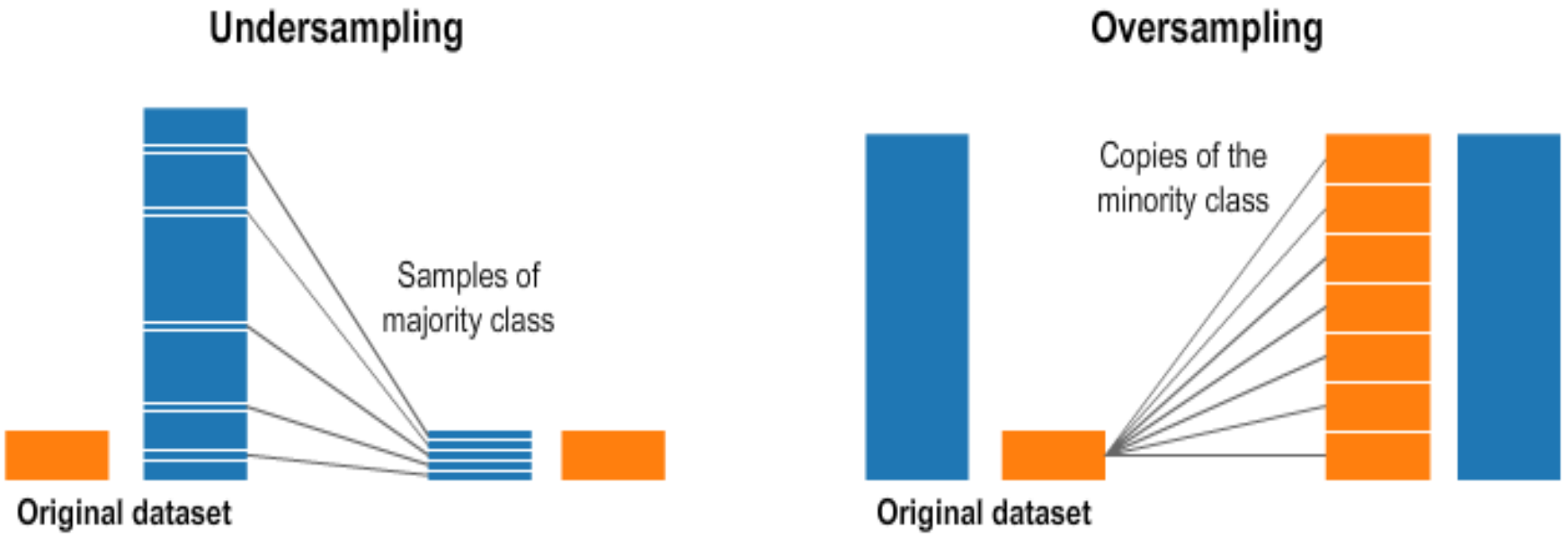
RESULTS ARE LOOKING GOOD

- WHY ?
  - Because the dataset is purely dominated with one Label
  - It's easy to predict that dominated label.

Class Imbalance

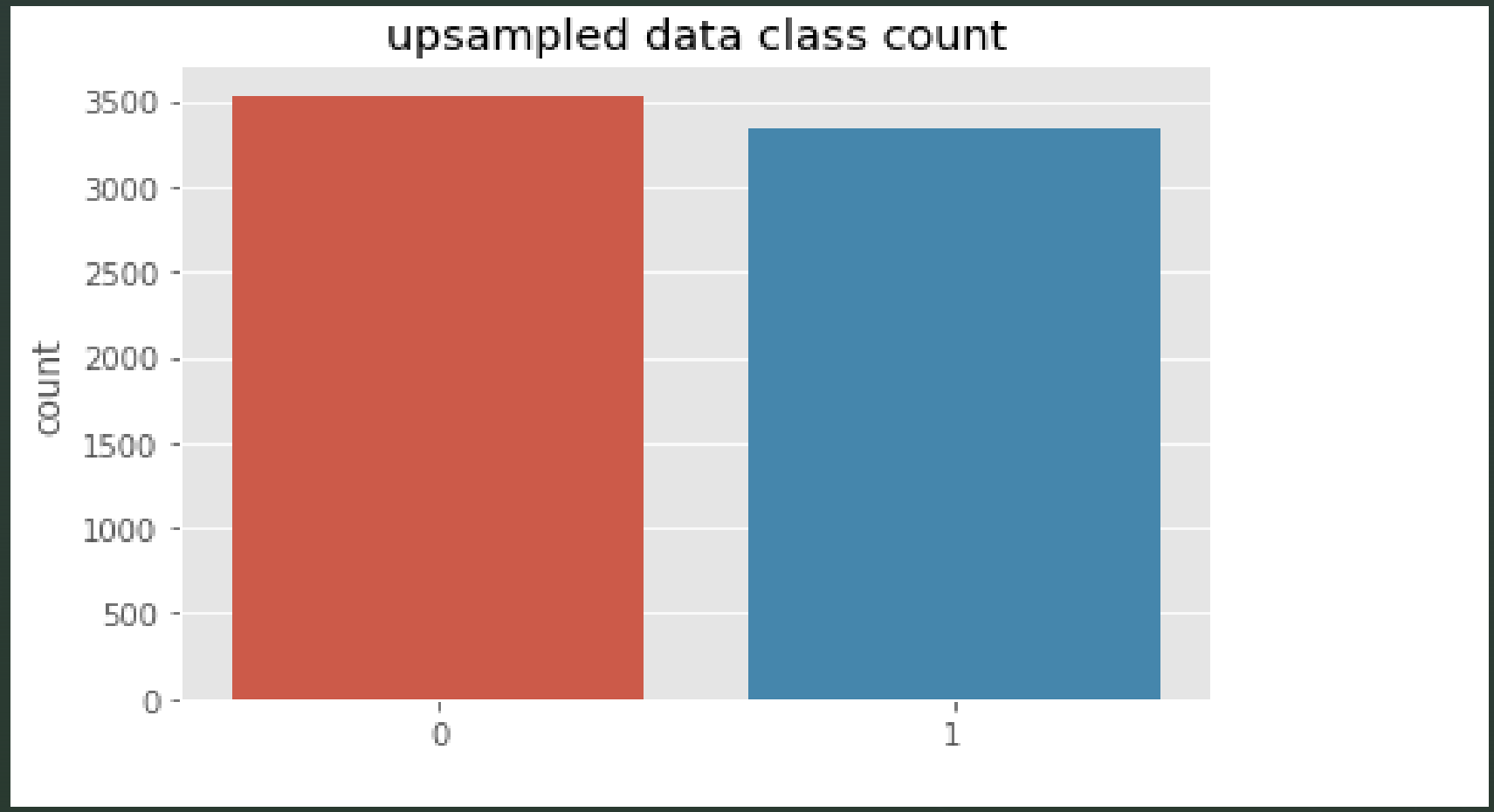


How to mitigate this

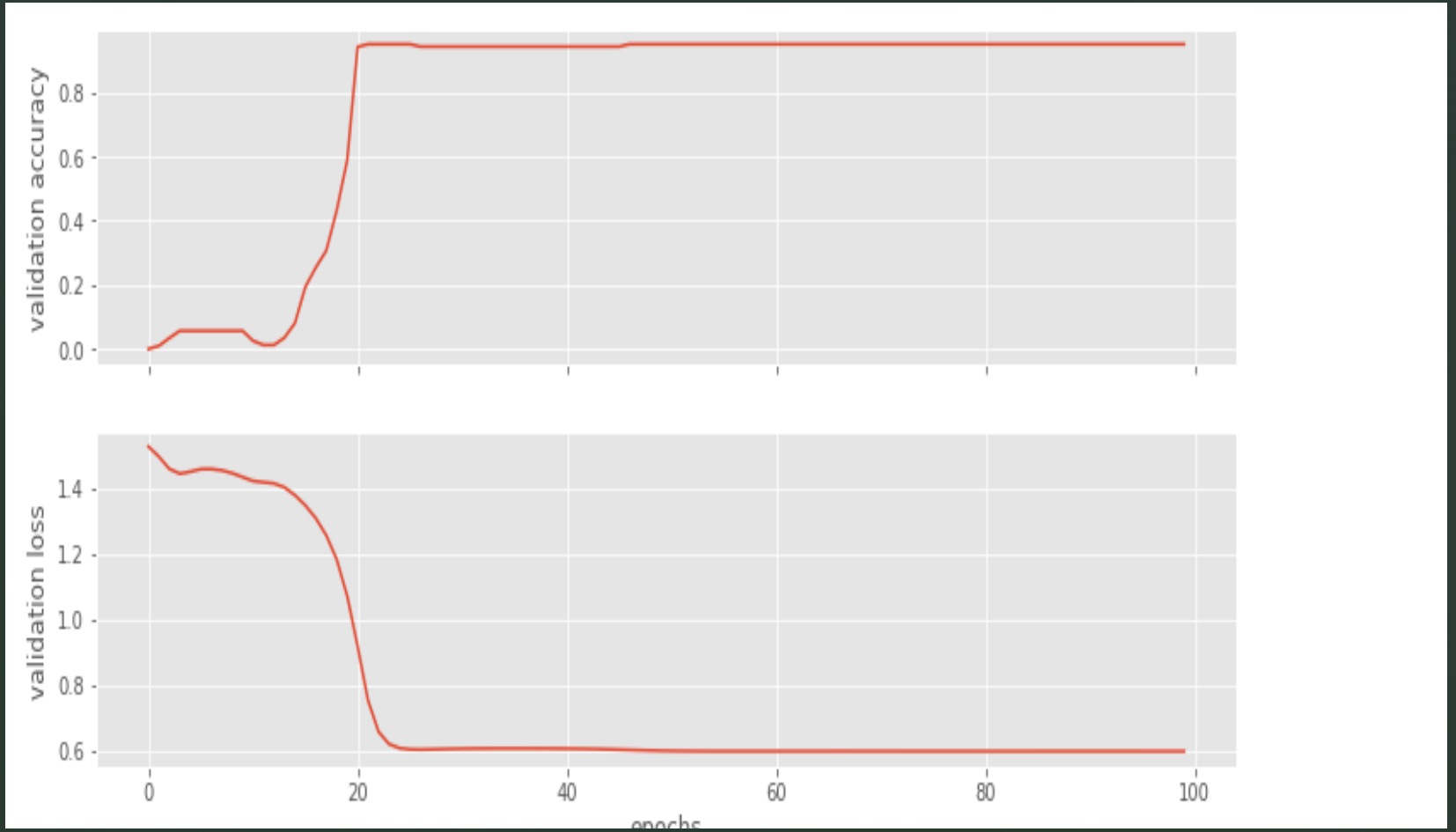


Mitigations of Class Misbalancing

- Dimensionality Reduction and Clustering
- Python Imbalanced Learn Module


upsampled data class count
count
3500
3000
2500
2000
1500
1000
500
0
0
1

Metrics


validation accuracy
0.0
0.2
0.4
0.6
0.8
validation loss
0.6
0.8
1.0
1.2
1.4
0
20
40
60
80
100
epochs

Metrics

```
array([[985,  28],
       [ 15,  32]])
```

```
Precision Score 0.8266051962781711
Recall Score 0.7591666666666667
```

FAIRNESS METRICS

```
Confusion matrix for Group-A [[213  10]
 [  1  11]]
Accuracy Score for Group-A 0.9531914893617022
```

```
Confusion matrix for Group-B [[109   2]
 [  0   5]]
Accuracy Score for Group-B 0.9827586206896551
```

Conclusion

- We found minimal fairness issues with Group-B

How to exclude bias in our words

- Data Processing is an important factor
- Choosing the best algorithms according to our data might also help

- Social Norm Bias

Research Papers

1. https://arxiv.org/pdf/2108.11056.pdf
Intensive Preprocessing of KDD Cup 99
2. https://arxiv.org/ftp/arxiv/papers/1805/1805.10458.pdf
3. A Fast Probing Detection Method using Hybrid Machine Learning Algorithms"
4. Evaluating Shallow and Deep Neural Networks for Network Intrusion Detection Systems in
Cyber Security
5. A Novel Statistical Analysis and Autoencoder Driven Intelligent Intrusion Detection
Approach

6. http://isyou.info/jisis/vol9/no4/jisis-2019-vol9-no4-01.pdf

7. Towards Intelligent Intrusion Detection Systems for Cloud Computing
8. Designing a Machine Learning Intrusion Detection System
9. http://cs229.stanford.edu/proj2017/final-reports/5230994.pdf
10. Machine Learning for Network Intrusion detetction System.
11. Improved Anomaly Detection Using Adversarially Learned Inference
12. A Virtual Machine Introspection Based Architecture for Intrusion

Thank you