CS 471: Statistical Methods in AI Project Presentation

Techniques for Network Intrusion Detection

Team Members
Murtuza Bohra (20172104), Rashmi KethiReddy (20172044),
Abhay Rawat (20172082), Nitin Ramrakhiyani (20172091)

Introduction - Problem

- Network Intrusions
- User Activities -> Features
- Track unusual activities. -> detect suspicious user activities
- Attacks fall into four main categories:
  - DOS: denial-of-service
  - R2L: unauthorized access from a remote machine
  - U2R: unauthorized access to local superuser
  - Probing: port scanning

Introduction - Data

- KDD99 10% dataset (~5 lakh records)
- Contains a total of 24 training attack types, divided into 5 categories
- Normal: 97,278 DoS: 391,458 probe: 4107 r2l: 1126 u2r: 52
- Includes 34 continuous and 7 categorical features
- Data Instance:
  0,tcp,http,SF,215,45076,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.0
  0,1.00,0.00,0.00,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,normal

Paper 1 (Feature Reduction) - Introduction

- Feature Reduction using various similarity measures
- Similarity between different dimensions taken as distance
- Similarity measures considered are:
  - Correlation Coefficient
  - Least Square Regression Error
  - Maximal Information Compression Index

Paper 1 (Proposed Method) - Contributions

Objective -> Classification

Idea for Feature Reduction : Across the Classes feature variance ⬆

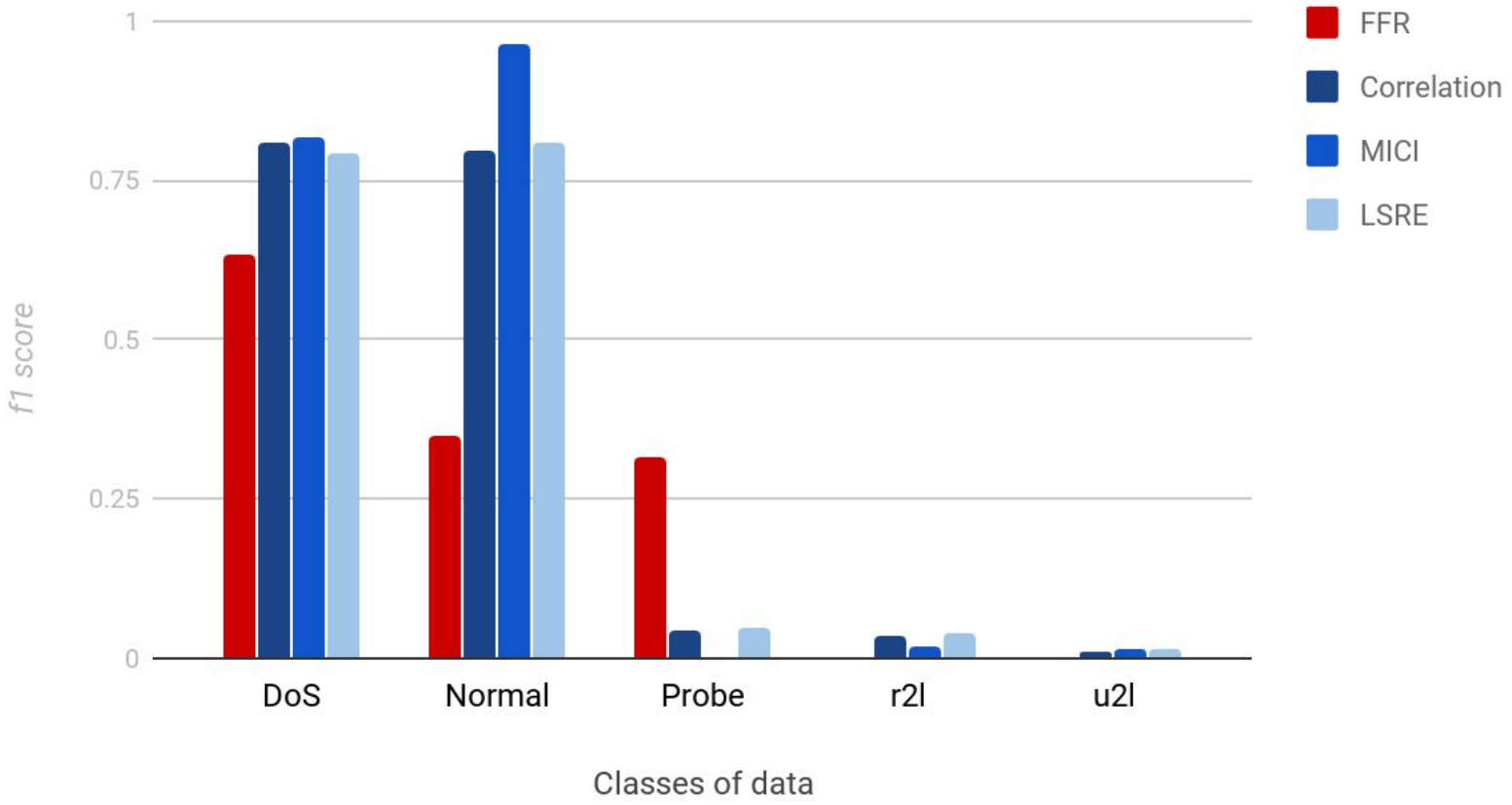
Paper 1 (Proposed Method) - Contributions (Cont.)

Challenges:

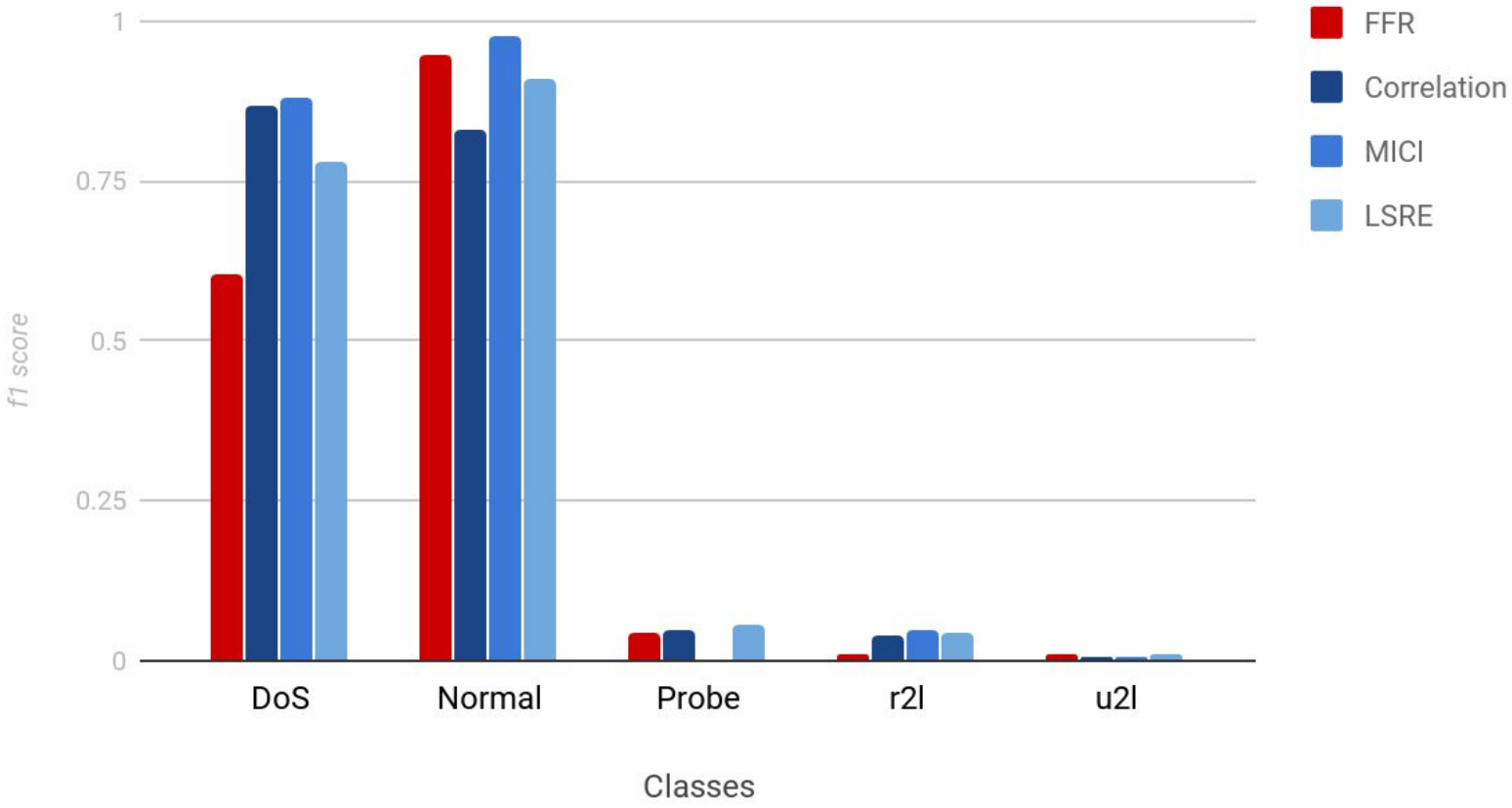
1) Variance for Categorical Features

2) Normalisation of features.

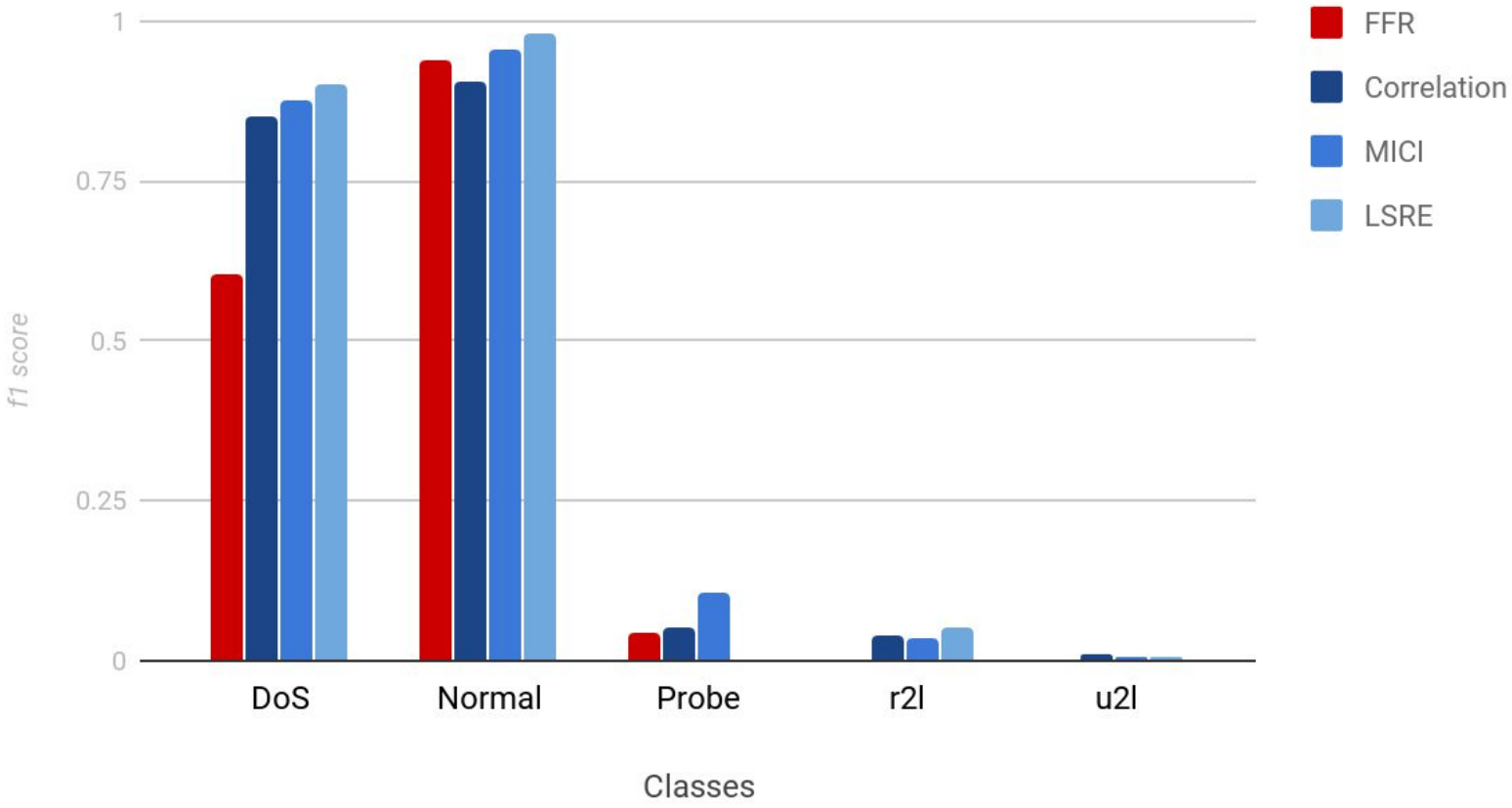
Feature Reduction on Bayesian (30 Features)



Feature Reduction on Bayesian (20 Features)



Feature Reduction on Bayesian (10 Features)



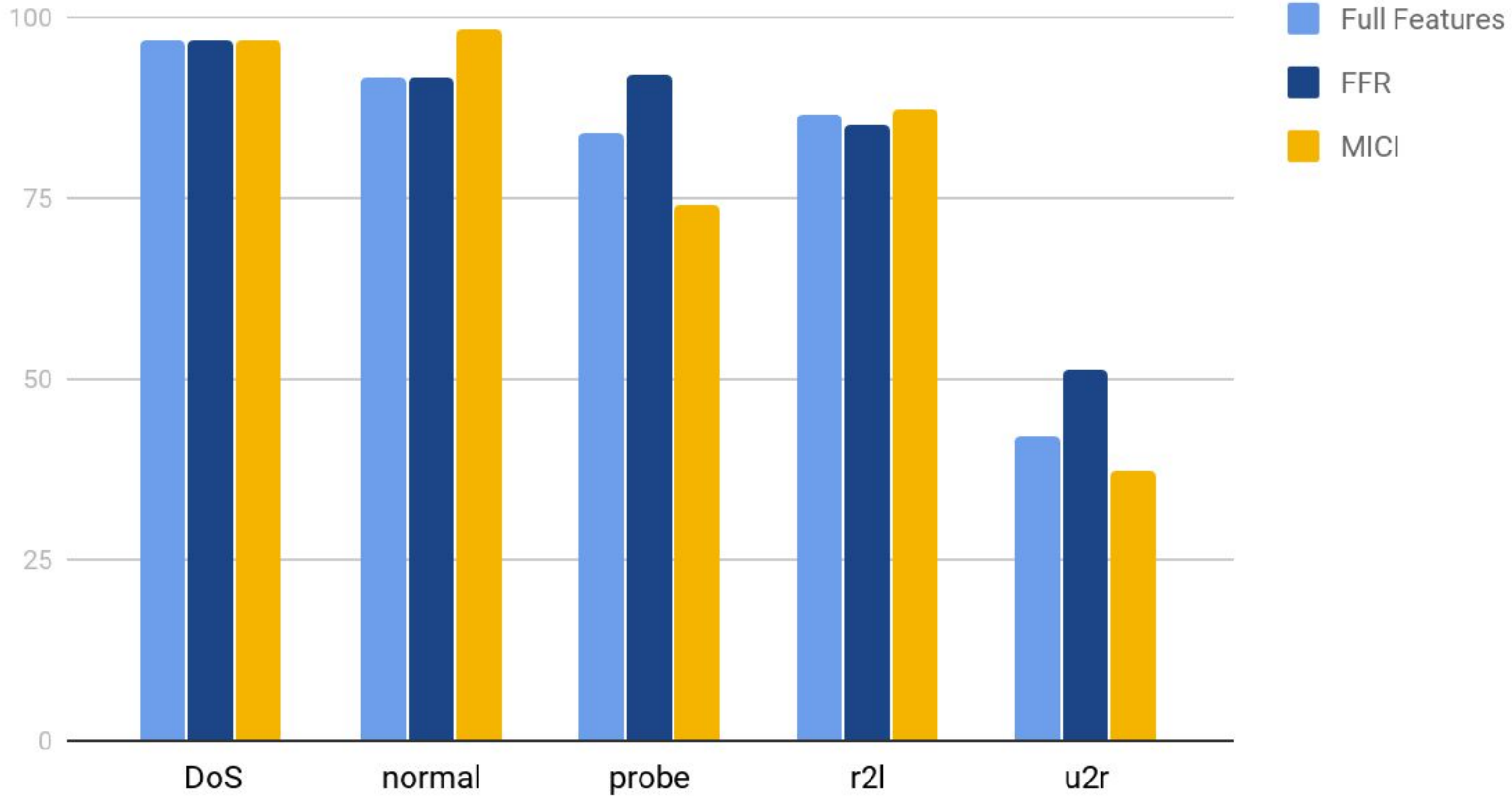
Paper 2 - Introduction

- Employed SVM and Neural Networks to classify intrusions
- Binary - attack vs normal
- Data used 14292 points with 7312 as training
- RBF kernel for SVM (Accuracy: 99.5%)
- Neural network with nodes as 41-40-40-1 (Accuracy: 99.25%)

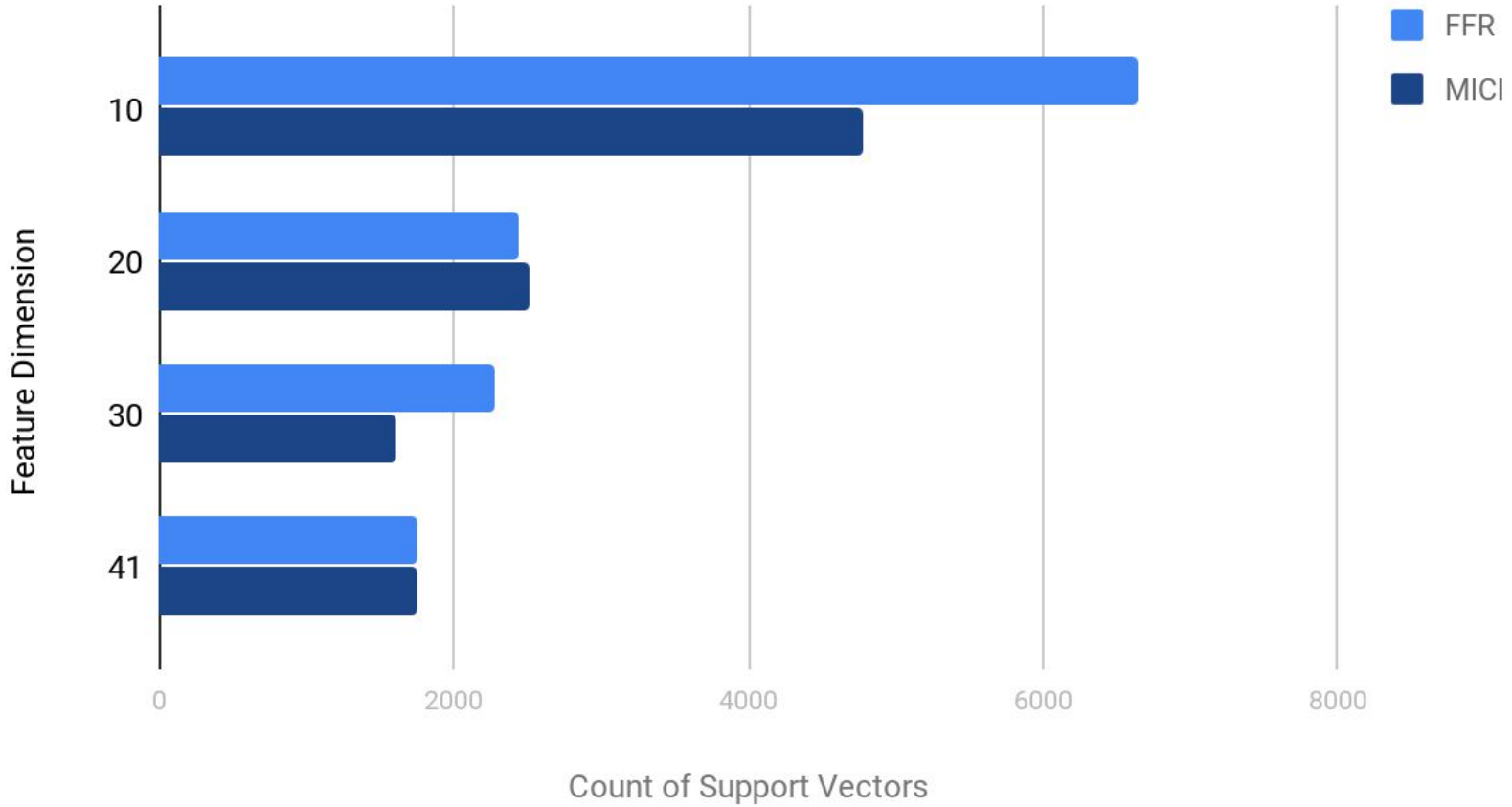
Paper 2 - Our Contributions

- Employed SVM and Neural Networks on full 10% data
- Tried both Binary and Multiclass (DoS, normal, …)
- RBF kernel for SVM with ‘ovo’ for multiclass
- Neural network with nodes
  - 41-40-40-1 for binary
  - 41-40-40-5 for multiclass
- Also employed with reduced feature sets of Paper 1’s techniques

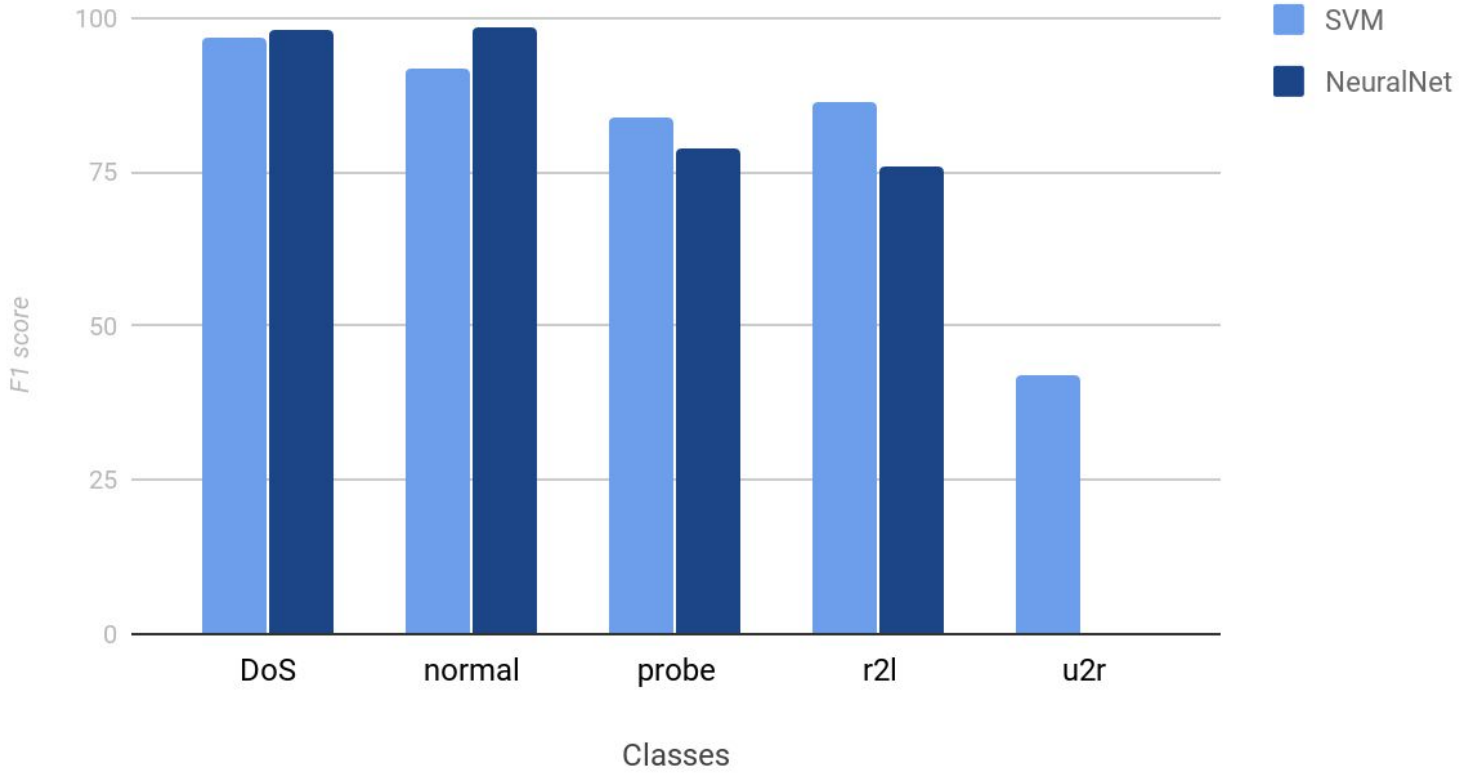
Effect of Reduced Features(20) on SVM



Count of Support Vectors Vs Feature Dimension



SVM vs NeuralNet (Classwise performance)



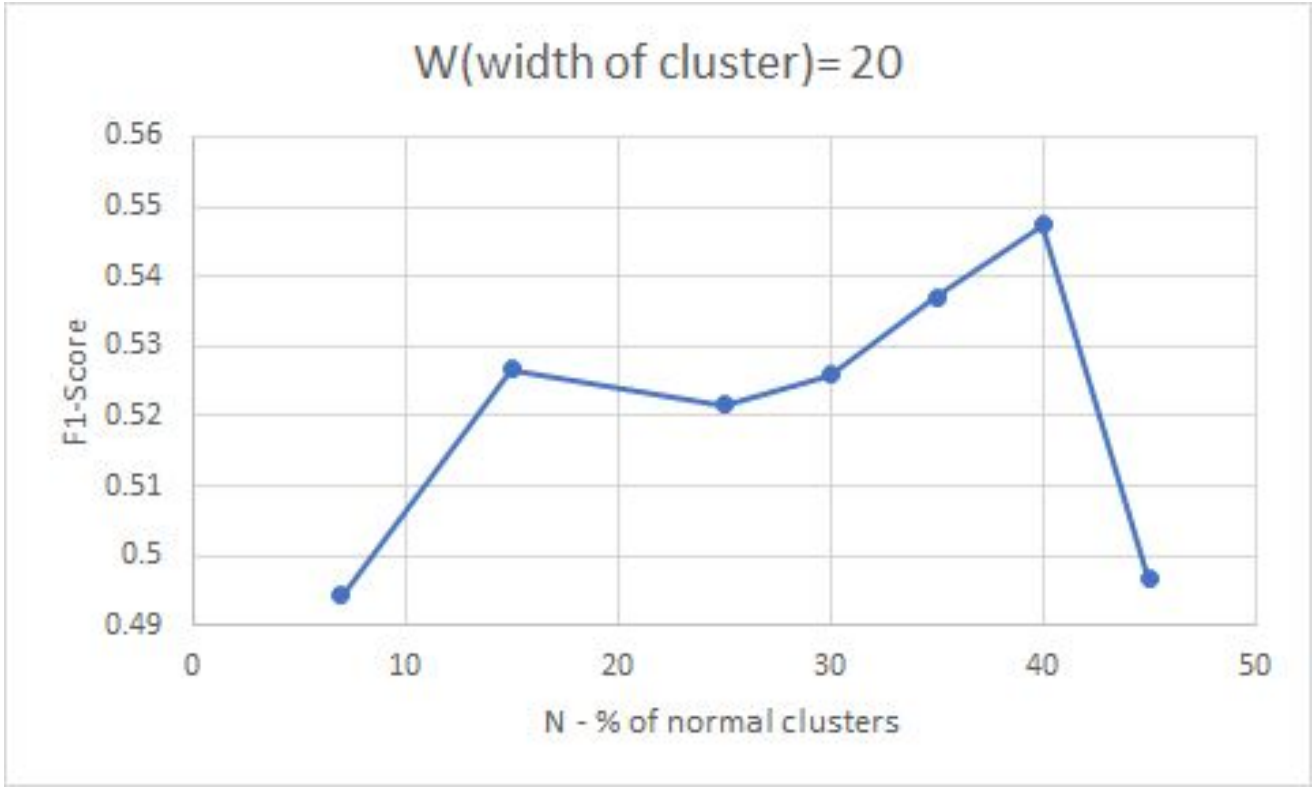
Paper 3 (Clustering) - Introduction

- Algorithms till now - Equal Distribution
- Real time scenario - Unequal distribution
- Anomaly Detection - Pure normal data (training) - tiresome & not accurate
- Cluster based unsupervised anomaly detection - Unlabelled , intrusions induced in data for training.
  - Normal data is more dense than the intrusion data
  - Normal instances vary qualitatively from intrusion

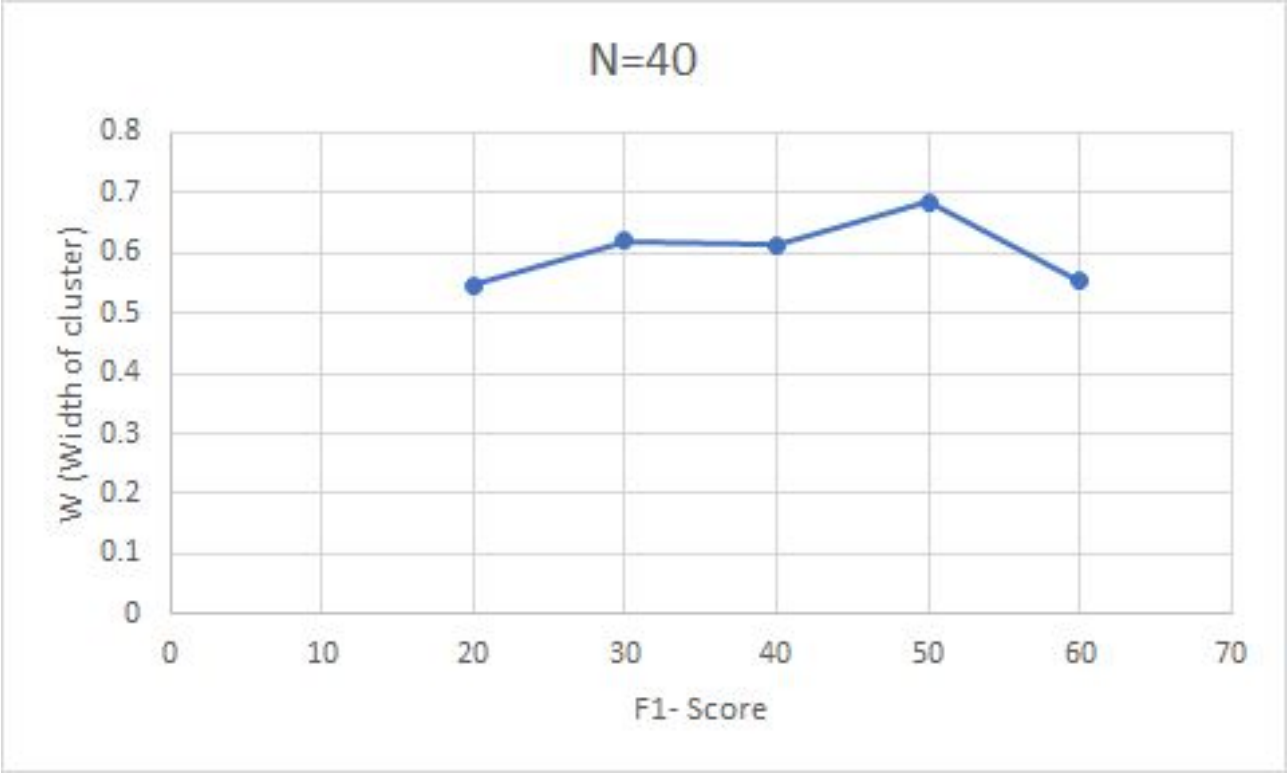
Paper 3 (Clustering) - Contribution

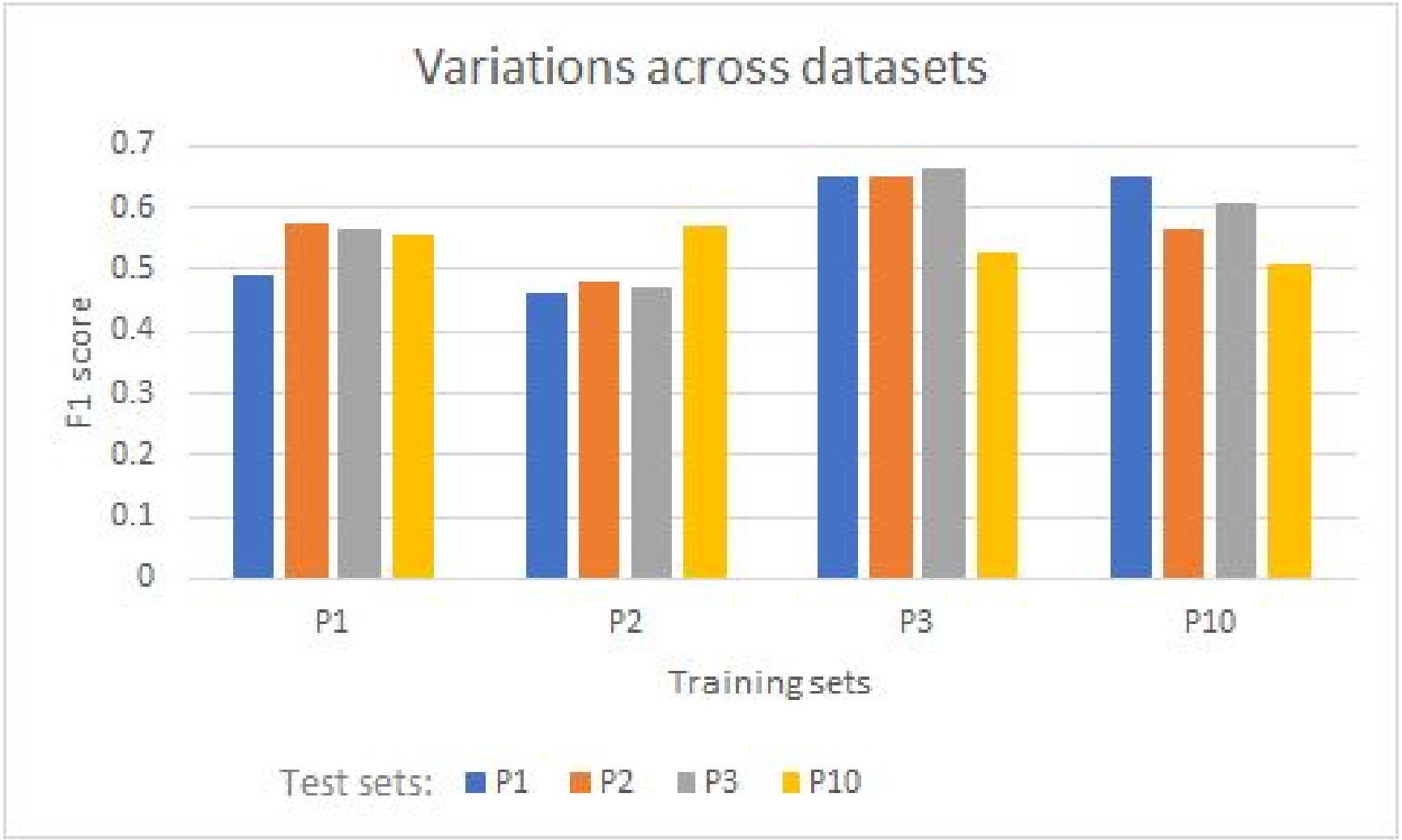
- Sampling KDD 10% Data - satisfy assumption 1
- Metrics - Euclidean distance (Normalize), Hamming (for categorical)
- Hyper parameters - W(Width of cluster), N(N% of clusters as normal)
- Labelling - Most populated clusters as normal
- Testing - Closest clusters class
- Clustered with reduced features of paper-1

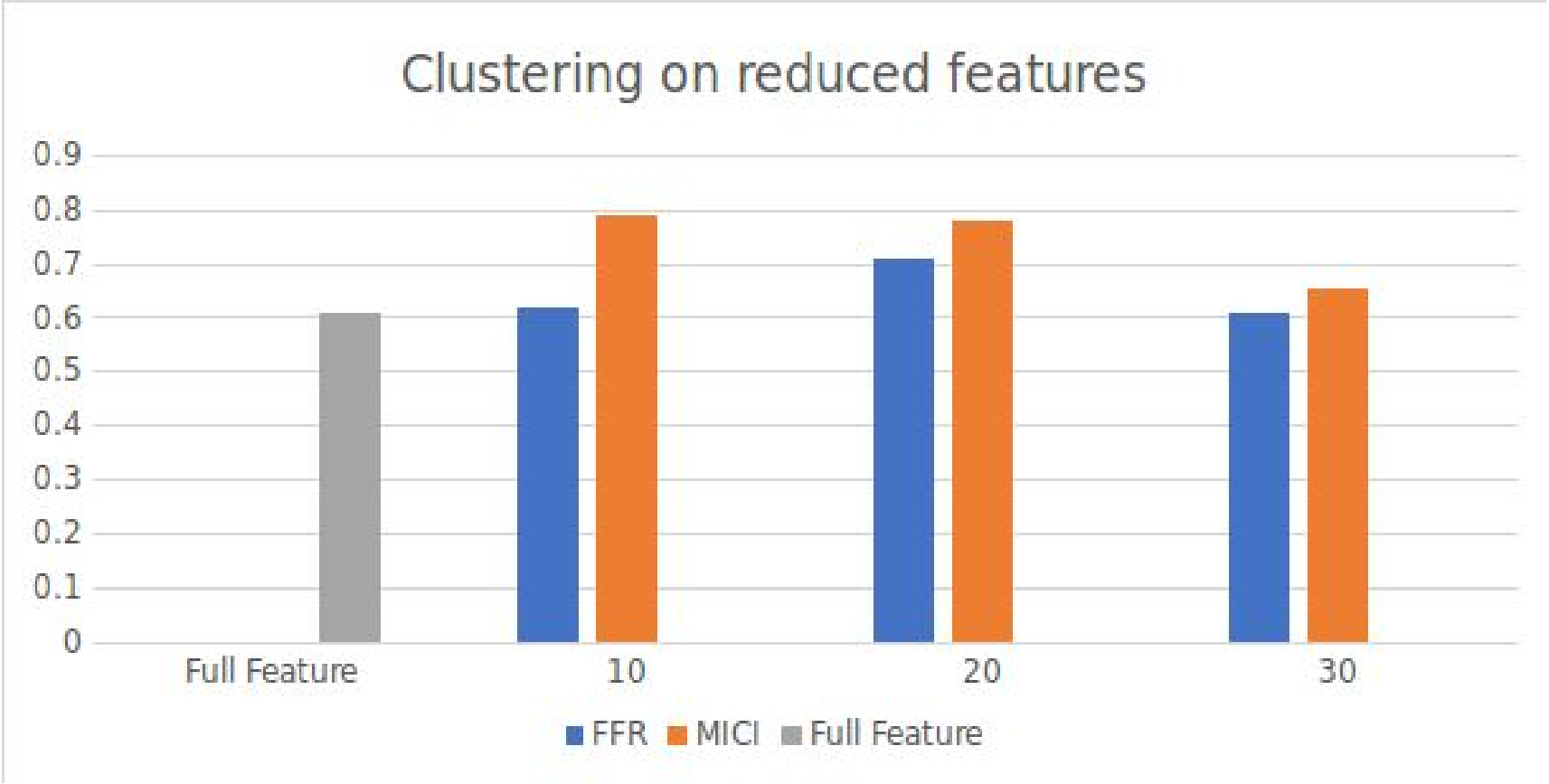
Fixing W and find N



Fixing N and finding W




Variations across datasets
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
F1 score
P1
P2
P3
P10
Training sets
Test sets:
P1
P2
P3
P10


Clustering on reduced features
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Full Feature
10
20
30
FFR
MICI
Full Feature

More...

- Probability based idea for variance of categorical features.
  - Notion of mean & variance.
  - How Probability will help?
  - Why it is intuitively better than one-hot encoding and numeric representation?

Thank you!

Questions?