Prediction and Anomaly Detection for Enterprise Access Control

Andre D'Souza

Supervised by Dr. Yuri Lawryshyn

April 10th, 2019

Agenda

1. Best Practices for Access Control
2. Objective Statement
3. Recommendation Systems for Prediction
4. PyOD Toolkit for Anomalies
5. Methodology & Data
6. Results
7. Conclusions & Recommendations

Best Practices for Access Control

Role Based Access Control (RBAC)

- Preferred network access solution in organizations over 500 members [1]
- Predefines roles defined with permission lists, and maps users role(s) based on access needs

Gap

- Role explosion caused by overly specific roles [1]
- Set covering issues when insufficient roles defined [1]
  - Under-extension: access omissions cause downtime
  - Over-extension: excess permissions increase exposure

Objective Statement

The objective of this research is to detect anomalies in existent access control configurations, and apply recommendation systems to predict future access control mapping and maximize enterprise network security.

Significance

- RBAC has undergone limited development[1]
- Minimize data security vulnerabilities
- Dynamic access control allocation
- Reduce network administration costs

Recommendation Systems for Prediction

Types of Recommendation Systems

- Collaborative filtering groups user patterns [2]
- Item-item stabilizes data matrices [3]
- Content based groups item and user attributes [2]
    - Attribute Based Access Control

Considerations

- Success determined by: a) Accuracy, b) User Satisfaction, c) Provider Satisfaction [2]
- Drawbacks due to cold start and data sparsity [2]

Anomaly Detection & Kaggle Challenge

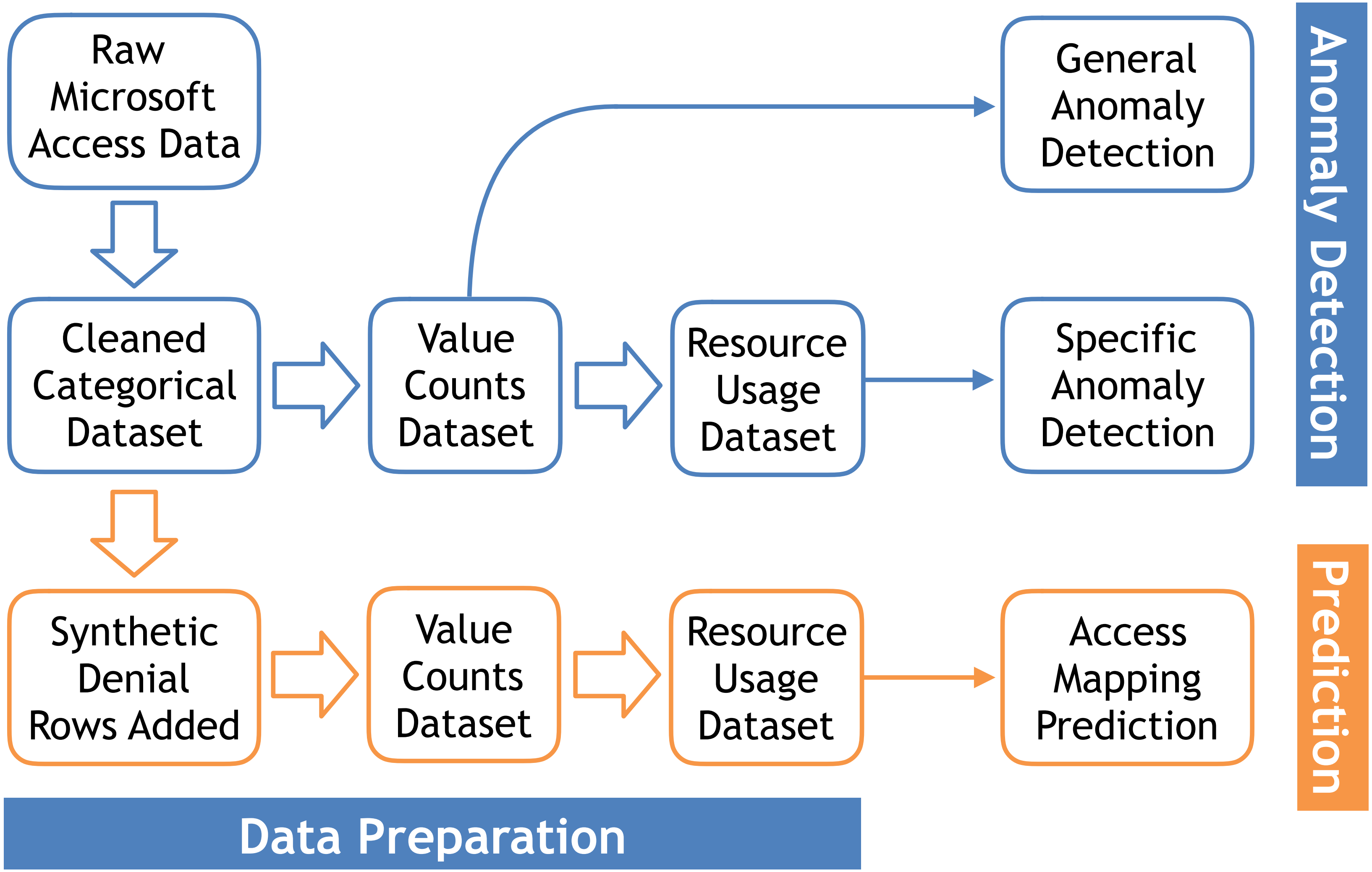
Tools for Anomaly Detection

- PyOD toolkit for Python selected [4]
- Three classifiers applied:
  - Isolation Forest
  - Histogram Based Outlier Selection
  - Cluster Based Local Outlier Factor

Kaggle Amazon Access Challenge

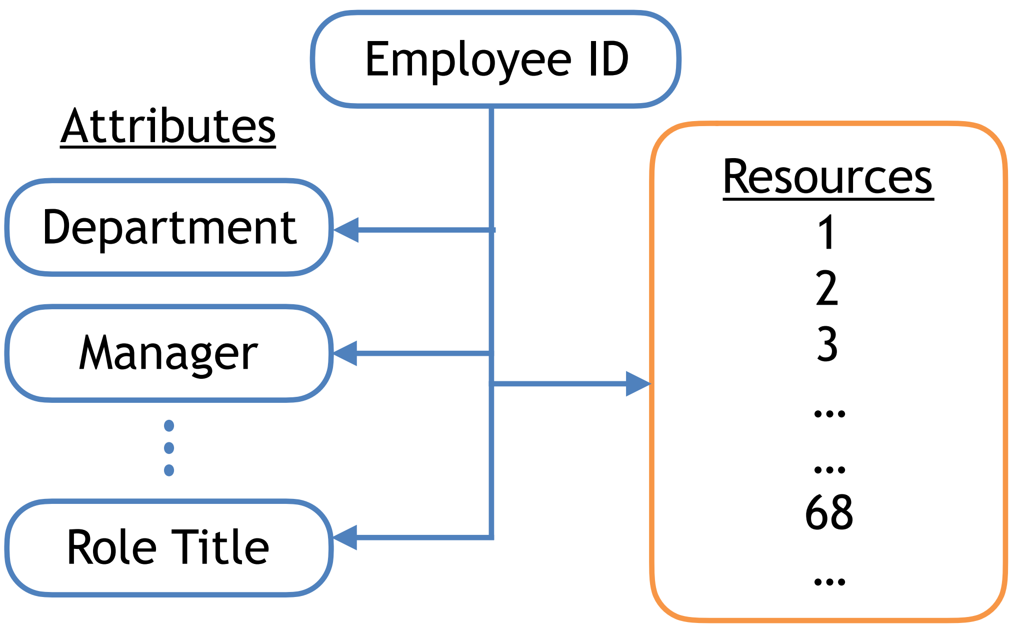
- Predict approval for employee resource access requests
- Data included resource requested & employee attributes

Methodology Overview



Microsoft Access Data

- MS Access used to export permission mapping into XML
- Parsed into data frame structured similar to Kaggle data
- Made each data frame row one employee resource pairing




```
<searchResultEntry dn="5335de625c5
   <attr name="manager">
      <value>87c8ff02614621d46a914
   </attr>
   <attr name="bufugu">
      <value>39a5a276d3830e76d7685
   </attr>
   <attr name="business">
      <value>698a9d279fd2ee19e6acf
   </attr>
   <attr name="ccode">
      <value>275f24097e3fa70b6466e
   </attr>
   <attr name="resource">
      <value>15799d193ec684e362291
      <value>007d8dcaf543bf3d77dba
      <value>39184e6db22e685f5573c
```

Training Dataset

- Each row represents a unique employee resource pairing
- Ex. 956 employees, each has an average of 68 resources
  - Association of 1.47% per resource (1/68)

	Resource	Manager	Depart.	Title	Bufugu	Business	Ccode	DN
Data Frame A: Cleaned RBC Dataset								
Count	66288	66288	66288	66288	66288	66288	66288	66288
Unique	7982	684	491	730	75	211	10	956
Data Frame B: Value Counts								
Median	104	112	184	91	5804	796	35248	68
Max	956	660	1983	942	11591	3571	47169	260
Data Frame C: Resource Usage Percentages								
Median	–	1.37%	1.27%	1.37%	0.66%	0.91%	0.21%	1.47%
Max	–	3.03%	2.86%	2.86%	2.56%	2.70%	2.08%	3.03%

Methodology: Prediction

Synthetic Data

- Synthesized 4k resource request denials
  - All 66.3k rows from MS Access are approvals
  - Appended employee-resource pairings absent from original dataset

Classification

- Used 80/20 training & test split
- Applied Random Forest, Extra Trees and Gradient Boosting for approval prediction
  - Preliminary recommendation system

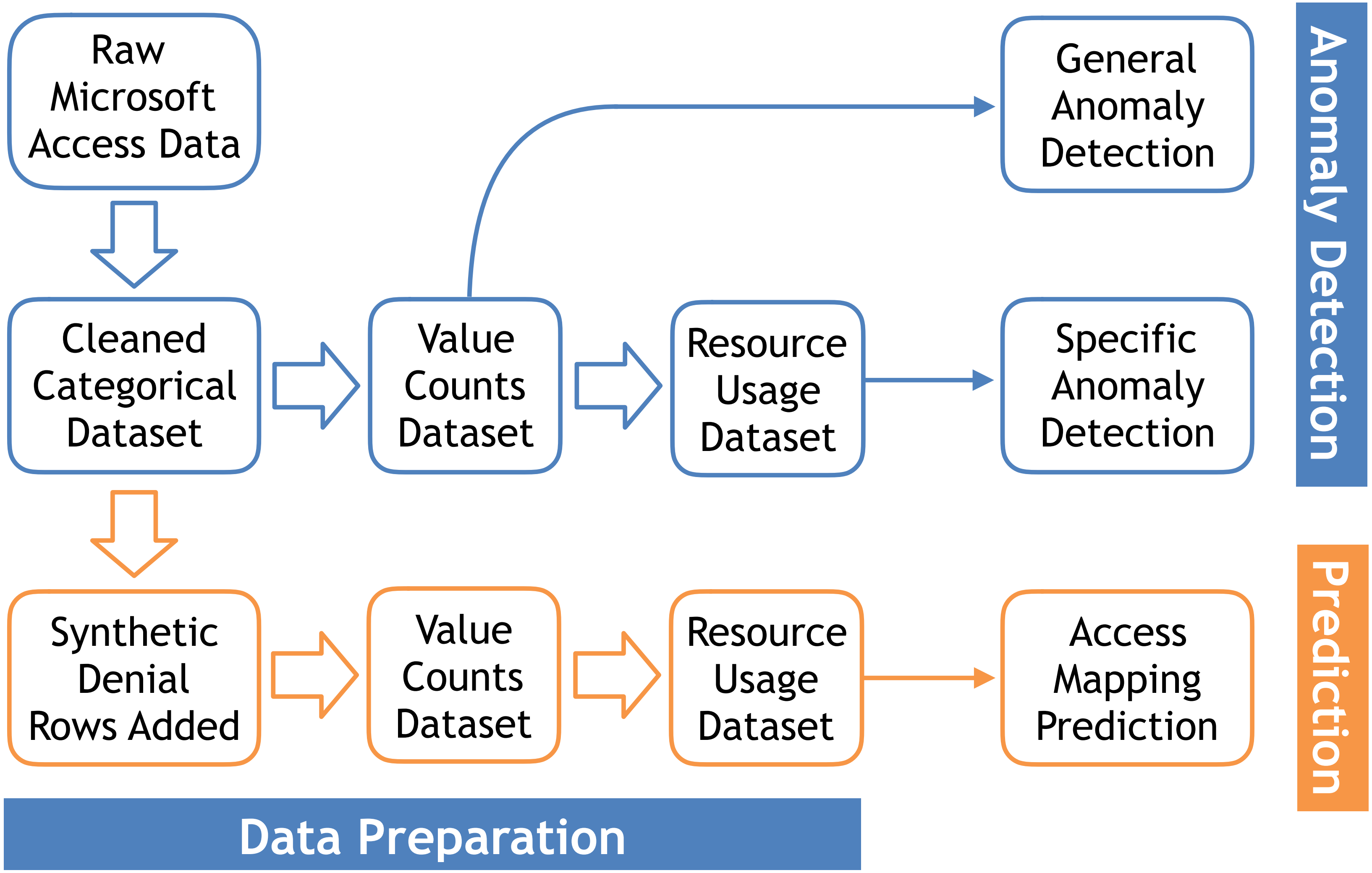
Results: Prediction

- Predicting access control mapping
- To achieve 94.3% approval rate, synthesized 4k denials

Probability of Resource Request Approval:

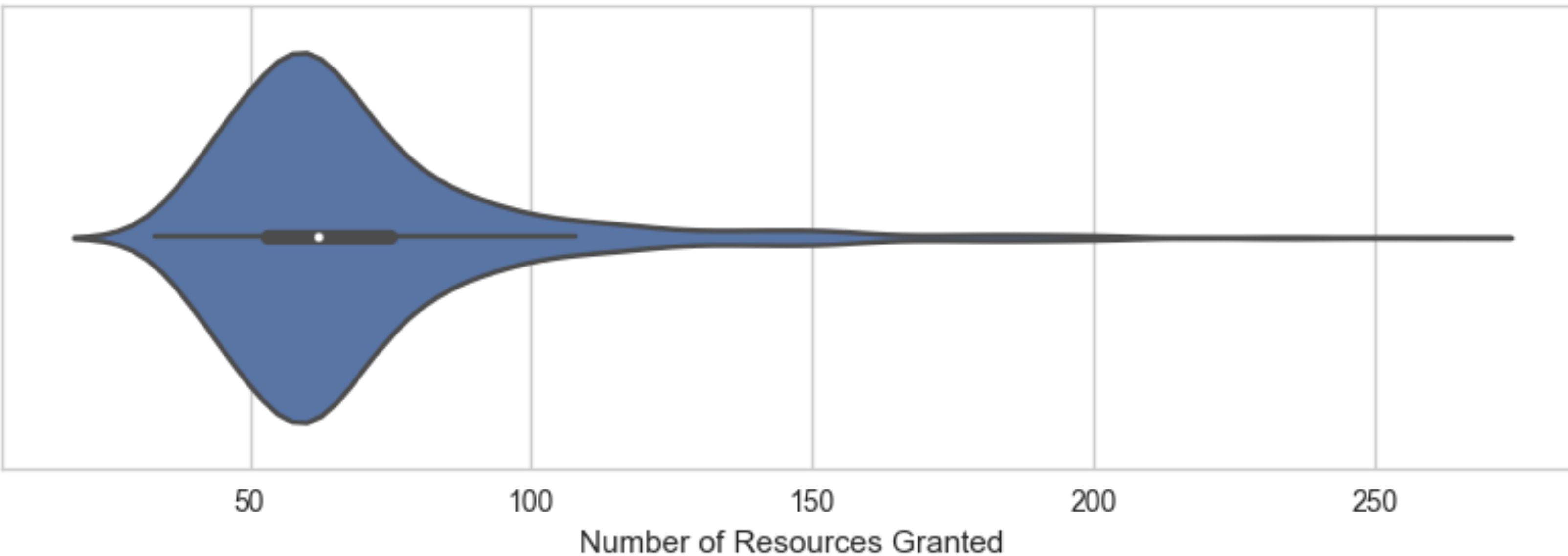
Statistic	Random Forest	Extra Trees	Gradient Boosting
Mean	92.299%	90.974%	97.532%
Minimum	7.091%	4.171%	0.013%
25th Percentile	89.633%	89.215%	99.668%
75th Percentile	98.086%	95.377%	99.696%
Maximum	100.000%	100.000%	99.699%

Methodology Overview



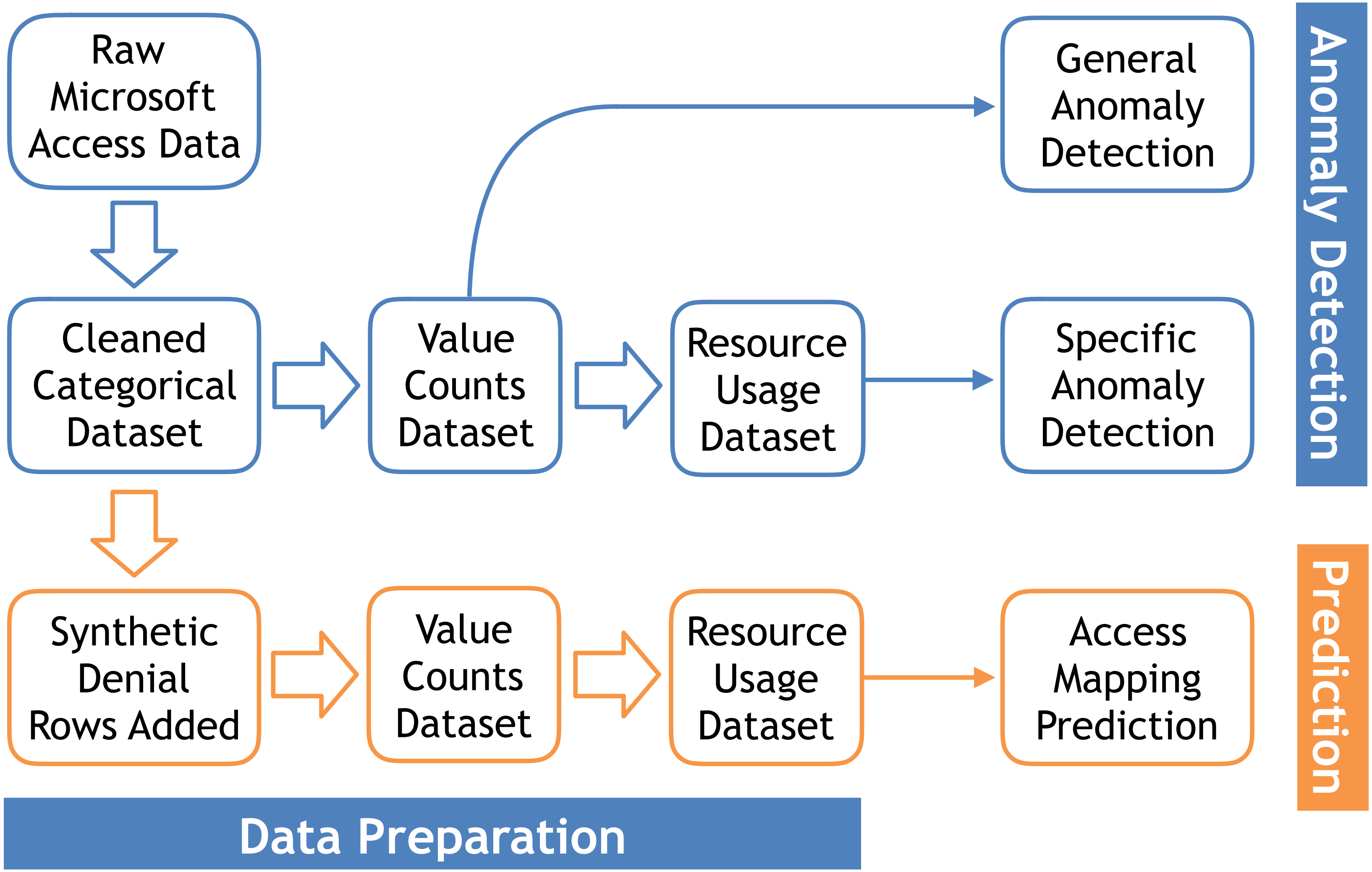
Results: Anomalies

Employees with high access count:

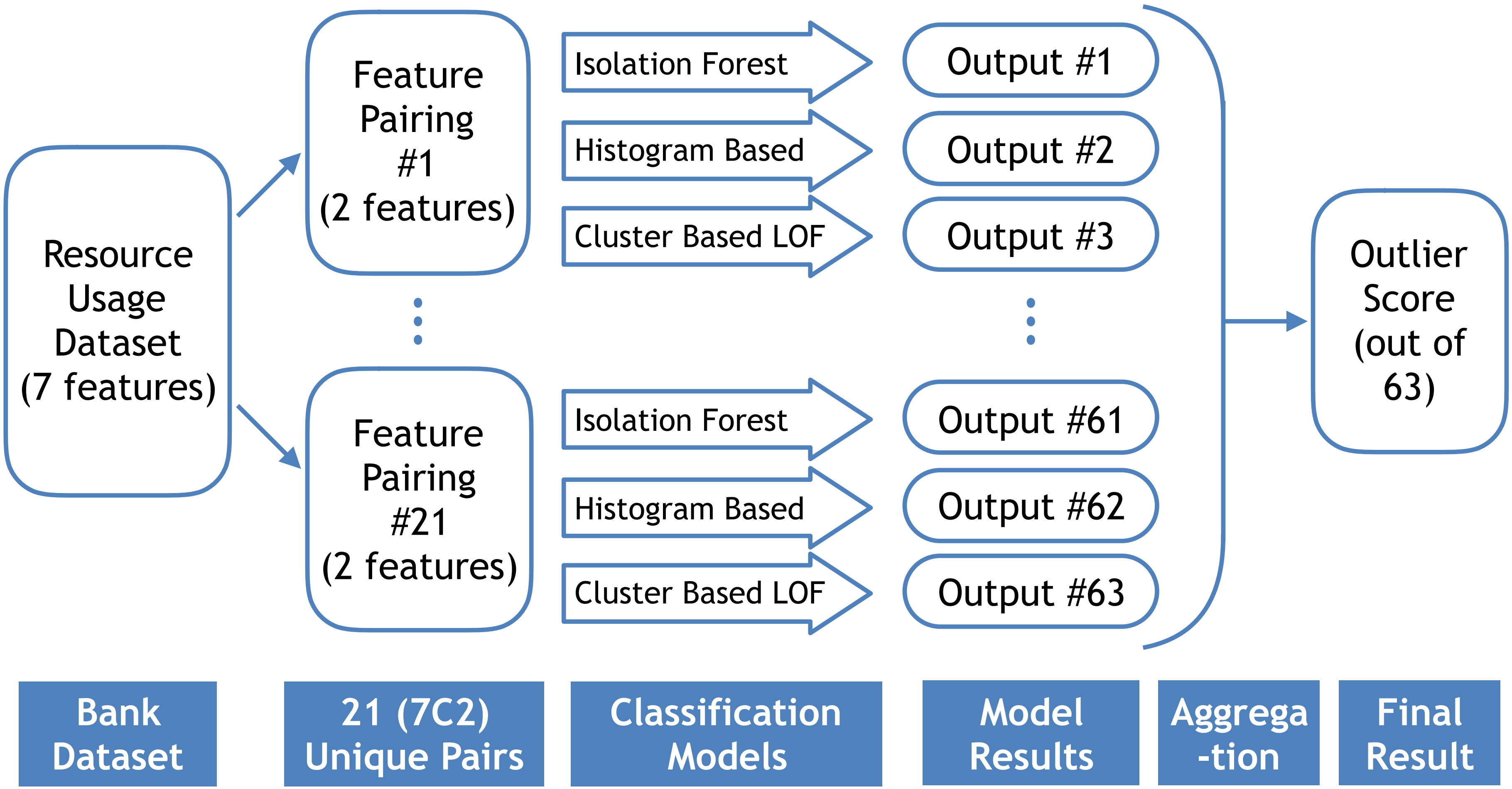


Resources Granted Threshold	25	50	100	150	200
Frequency of Employees	956	770	95	27	5

Methodology Overview

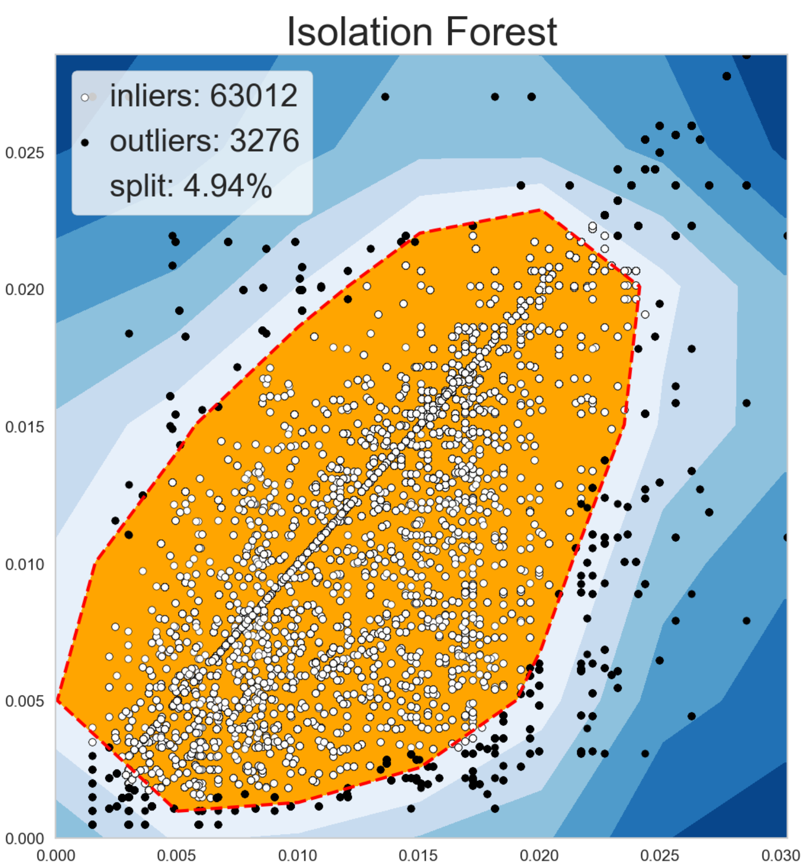


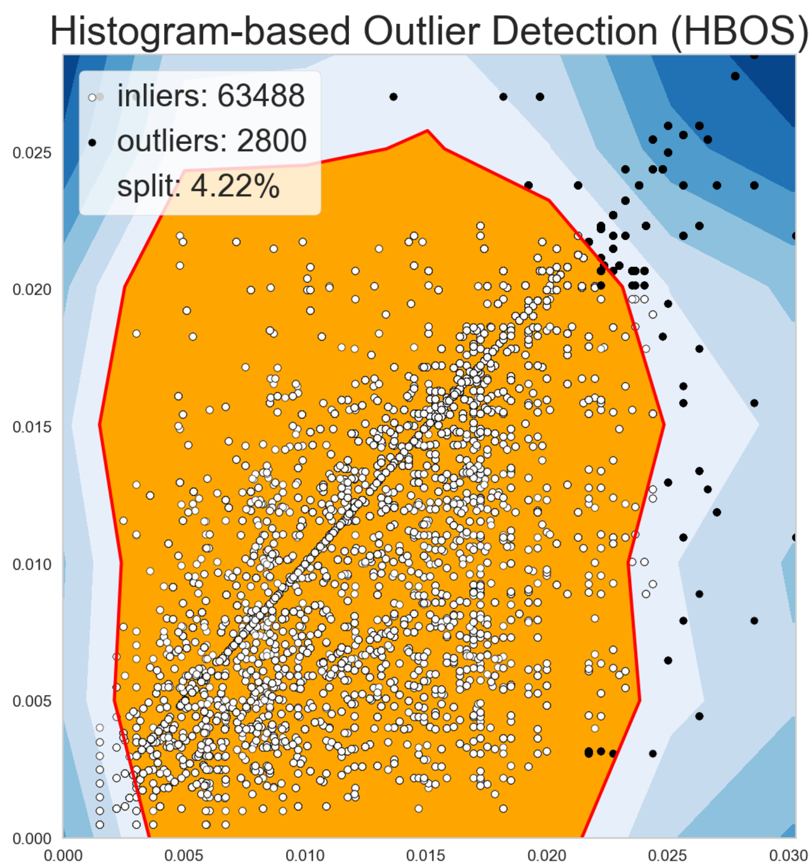
Methodology: Anomalies



Results: Anomalies

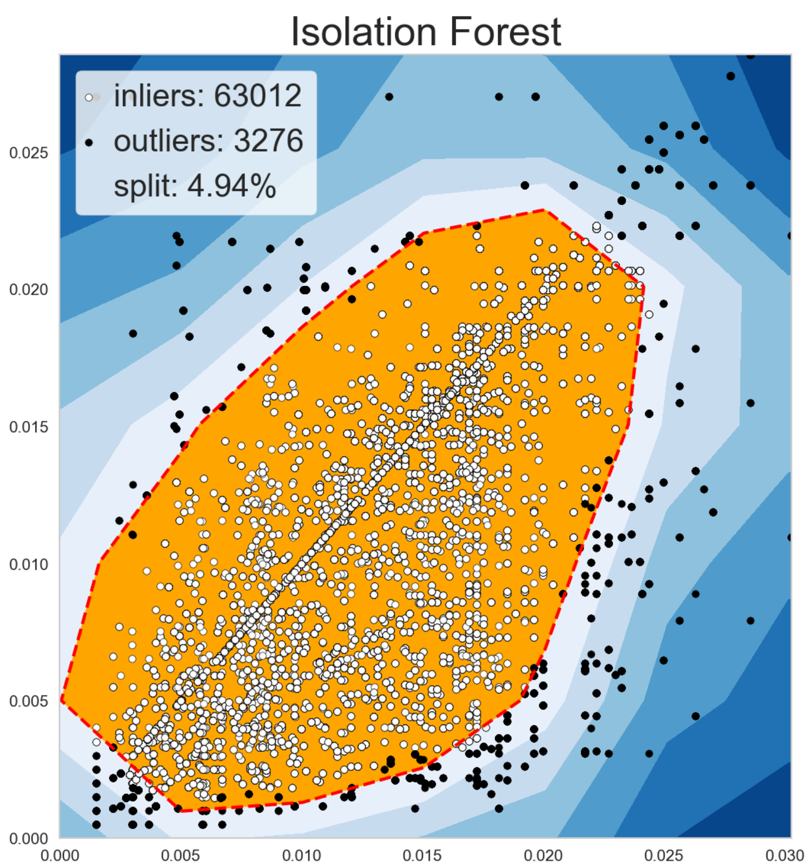
Manager vs Department Resource Usage Percentages

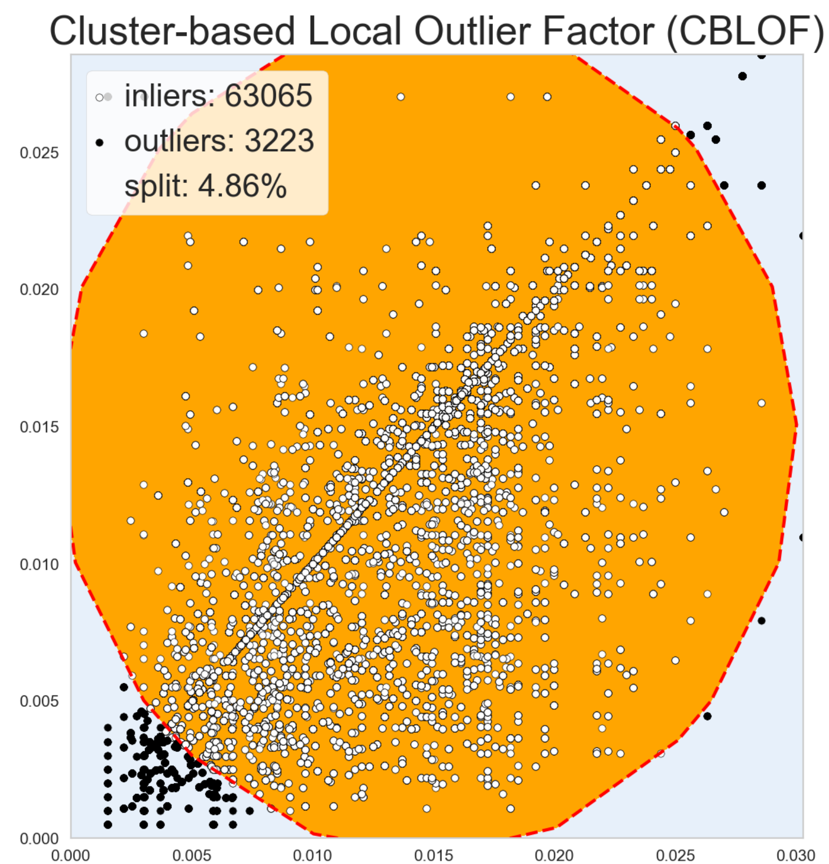




Results: Anomalies

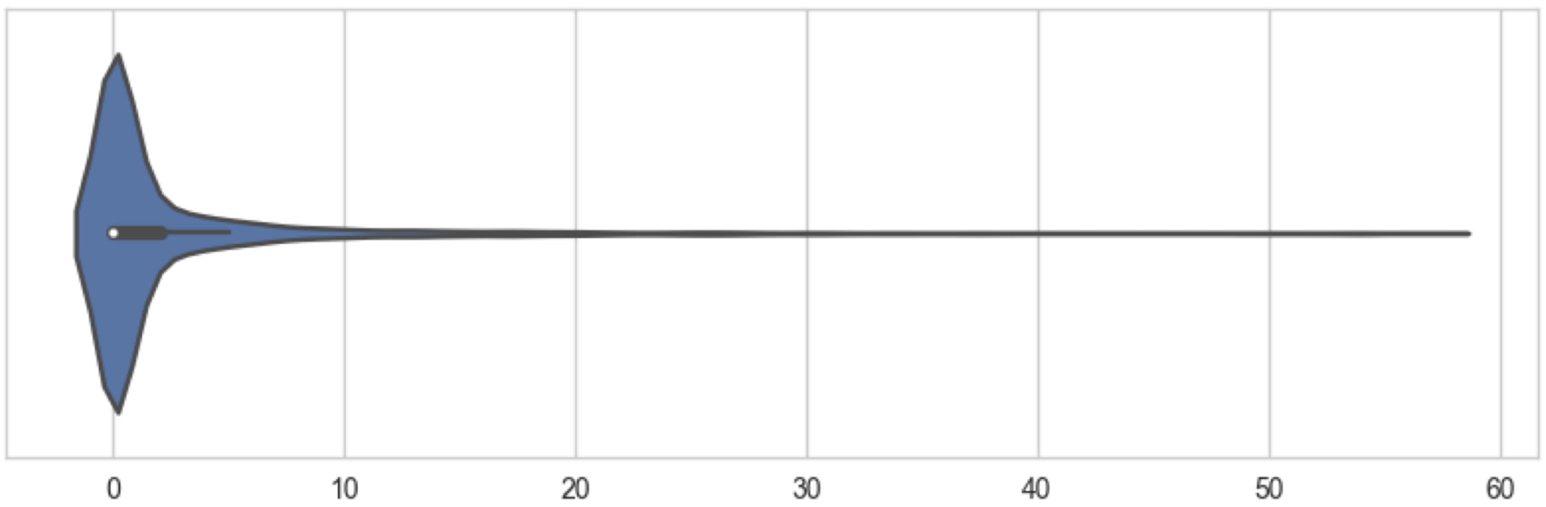
Manager vs Department Resource Usage Percentages





Results: Anomalies

Outlier Score: aggregates 63 classifications



Score Threshold	Outlier Count	Inlier Count	% Outliers
0	25820	40468	38.95
5	9546	56742	14.4
10	5459	60829	8.24
20	2491	63797	3.76
30	1300	64988	1.96
40	718	65570	1.08
50	312	65976	0.47

Results: Anomalies

- Classified full dimension dataset
- 5% corresponds to an Outlier Score of 16-17
  - Encompasses around 144 employees
- Can use to control scope of review

Classification Algorithm	Outlier Count	Inlier Count	% Outliers
Isolation Forest	3314	62974	5.00
Histogram-base Outlier Detection	3313	62975	5.00
Cluster-based Local Outlier Factor	3311	62977	4.99

Conclusion

Access Control Prediction

- Recommendation systems a viable access control method
- Requires testing with unsynthesized denials before implementation

Anomaly Detection

- 5% of examined permissions granted are possible outliers
- Recommend reviewing by descending Outlier Score
- Cease review if low precision discovered

Future Work

Access Control Prediction

- Explore other data sources with request denials
- Apply Restricted Boltzmann Machines for better accuracy [5] and ensemble more predictors

Anomaly Detection

- Validate detected outliers
- Experiment with feature engineering & selection
- Apply alternative anomaly detection models

References

1. C. O'Connor and R. J. Loomis, "Economic Analysis of Role-Based Access Control: Final Report," Economic Analysis of Role-Based Access Control: Final Report | CSRC, 19-Dec-2010. [Online]. Available: https://csrc.nist.gov/publications/detail/white-paper/2010/12/19/economic-analysis-of-rbac- final-report/final.
2. J. Beel, S. Langer, M. Genzmehr, B. Gipp, C. Breitinger, and A. Nürnberger, "Research paper recommender system evaluation," Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation - RepSys 13, 2013.
3. Smith, B., & Linden, G. (2017). Two Decades of Recommender Systems at Amazon.com. IEEE Internet Computing, 21(3), 12-18. doi:10.1109/mic.2017.72
4. Zhao, Y., Nasrullah, Z. and Li, Z., 2019. PyOD: A Python Toolbox for Scalable Outlier Detection. arXiv preprint arXiv:1901.01588.
5. R. Salakhutdinov, A. Mnih, G. Hinton, "Restricted Boltzmann Machines for Collaborative Filtering", University of Toronto. November 2016.