# Binary Classification of Rare Events in Consumer Advertising Data
## Using Machine Learning Methods

**Jared Cook, Elizabeth Herman, Evan North, Katrina Petroske**
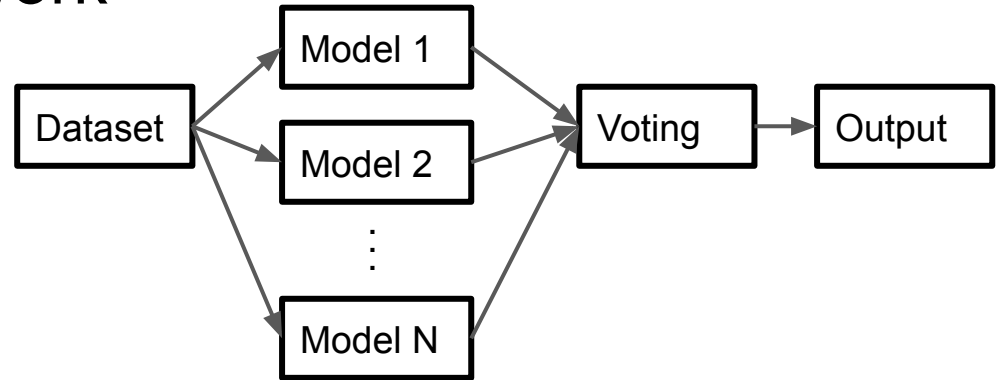
# Introduction and Motivation

- User shopping traffic from Valassis. Binary classification of "Converted" vs. "Not Converted".
- Conversion is a rare event (~1% of total).
- **Objective Question: Can conversion be predicted from a user's interest data?**
  - Accurate predictions of conversion leads to more profits.

- Approaches are Dimensionality reduction (PCA) and popular machine learning prediction techniques.

- 99% of variance explained in 453 singular vectors.
  - PCA space provides no clear linear classification.
  - Singular values provide no (useful) clear truncation point
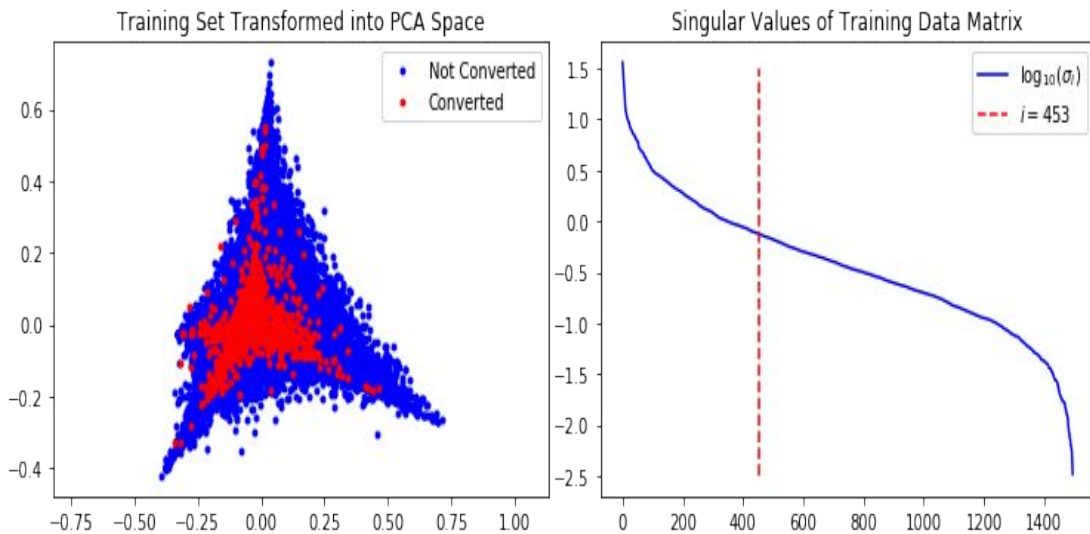
# Ensemble and Neural Network
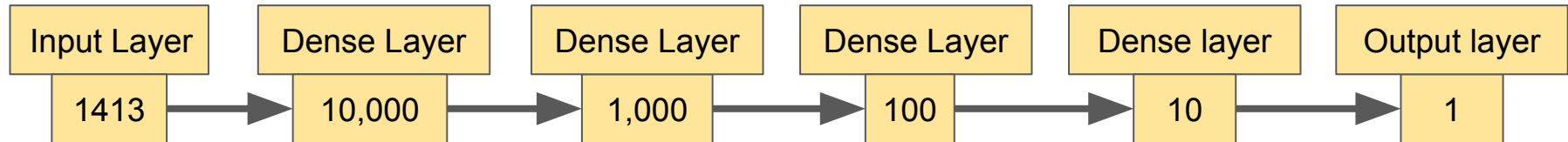
- Ensemble multiple methods
  - Naive Bayes
  - Random Forest
  - Logistic Regression
    - Cross-validation
    - ROC score
- Deep Neural Network (DNN)
  - Nonlinear dimensionality reduction
  - Fully connected, feed forward
  - Motivate autoencoder



| Ens | P | N |
|-----|---|---|
| T | 1 | 79386 |
| F | 2 | 619 |

| DNN | P | N |
|-----|---|---|
| T | 2 | 79385 |
| F | 3 | 618 |

| Input Layer | Dense Layer | Dense Layer | Dense Layer | Dense layer | Output layer |
|-------------|-------------|-------------|-------------|-------------|--------------|
| 1413 | 10,000 | 1,000 | 100 | 10 | 1 |

# Autoencoder

- Anomaly detection
  - Train on majority group
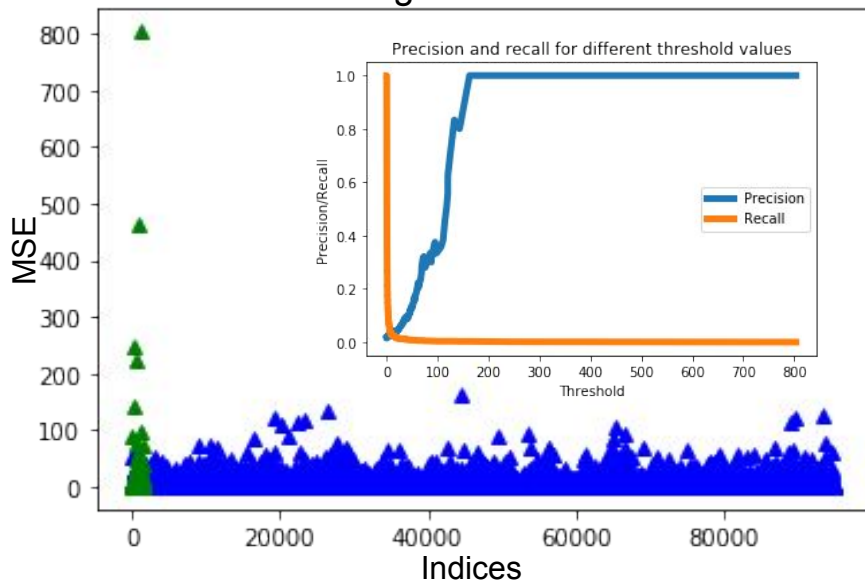  - Large MSE $\longrightarrow$ anomaly
  - Hyperparameter: threshold

| Input layer | Dense layer | Dense layer | Dense layer | Dense layer | Dense layer | Dense layer | Dense layer | Dense layer | Output layer |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| 1413 | 374 | 94 | 24 | 6 | 6 | 24 | 94 | 374 | 1413 |

Encoder                    Decoder

Training Data MSE

Precision and recall for different threshold values
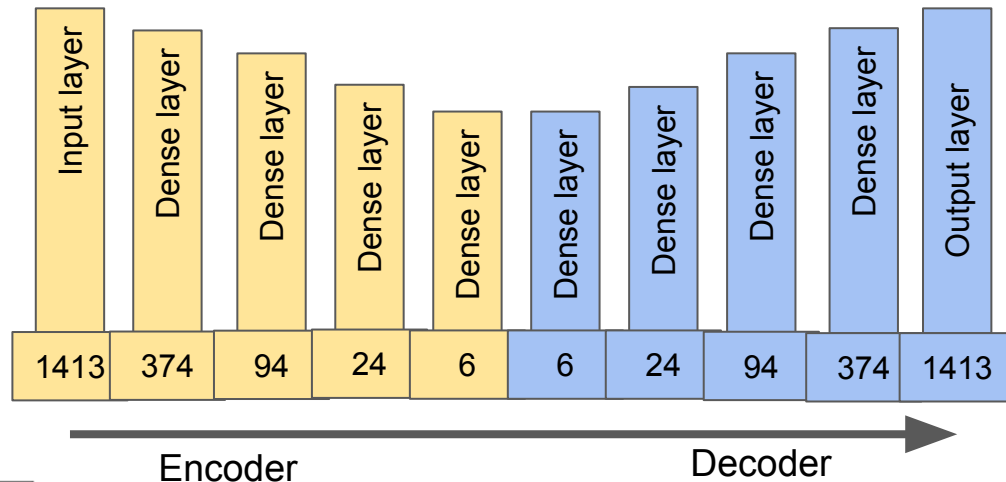
| AE 200 | P | N |
|:---:|:---:|:---:|
| T | 14 | 77622 |
| F | 1766 | 606 |

| AE 100 | P | N |
|:---:|:---:|:---:|
| T | 31 | 75591 |
| F | 3797 | 589 |

# Conclusions and Future Work

- Naive out of the box machine learning methods do not work well for rare event prediction
- Autoencoder produces best results using the confusion matrix metric
- Future work:
  - Hyperparameter optimization
  - Employ F1, F2, ROC metrics for performance and cross-validation
  - Other anomaly detection methods
    - Boosting methods (Gradient boosting/Adaboost)
    - SVM
  - Further exploration of data for feature selection
    - Non-linear dimension reduction
    - Can we target consumers that are more likely to convert?