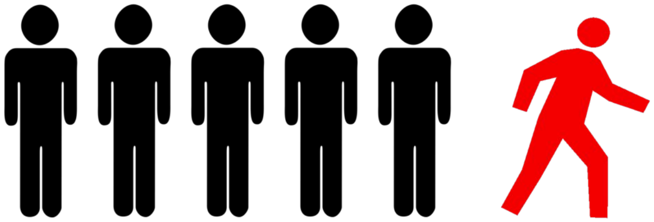
Proposed Churn Analysis Techniques

Definition:
For the purposes of Levi’s/Dockers Inc., we will define **churn** as the loss or outflow of members of the email subscription base by deliberate unsubscription.



Benefits of predicting churn:
- Attracting new clients costs 5 to 6 times more than retaining current customers (7).
- Prevent defection and increase ROMI (return on market investment) (6).
- Increase brand loyalty (8).

Common Approach:
- Train a “white box’ machine learning algorithm to understand underlying patterns/behaviors/interactions that point to relative high likelihood to churn.
- Target campaigns based on results.

Downsides of traditional model:

Customers can be categorized into 4 groups, **sure things**, **persuadable**, **negative impact** (Do-Not-Disturbs), and **lost causes**.

Churn when targeted	Yes	Do-Not-Disturbs	Lost Causes
	No	Sure Things	Persuadables
		No	Yes
		Churn when not targeted	

(1) Devriendt et. al.

Traditional approaches classify those likely to churn indiscriminately, resulting in outreach campaigns that waste money attempting to retain customers who will churn no matter what, and negatively impacting churn likelihood by marketing to customers who are averse to outreach.

Proposed Approaches: Deep Learning vs. Uplift model

- *Uplift modeling* can estimate the net effect of an outreach campaign and thus provide metrics for differentiating between categories. Customers with a high net treatment effect can be targeted, and others avoided. Using an uplift model was found to outperform predictive models and lead to improved profitability of retention campaigns when modeled on financial institution data **(1)**.

- *A deep learning model* can be used as a vector embedding model to visualize the relationship among churning and loyal customers. It can also differentiate persuadables in vector space **(2)**.

Uplift Model:
The goal of uplift modeling is to understand the change in likelihood of churn if a customer is targeted with a retention campaign.

Definition of Uplift:
The difference in probability of churn assuming no treatment and assuming treatment. The goal being to estimate the net effect on churn after applying a treatment.

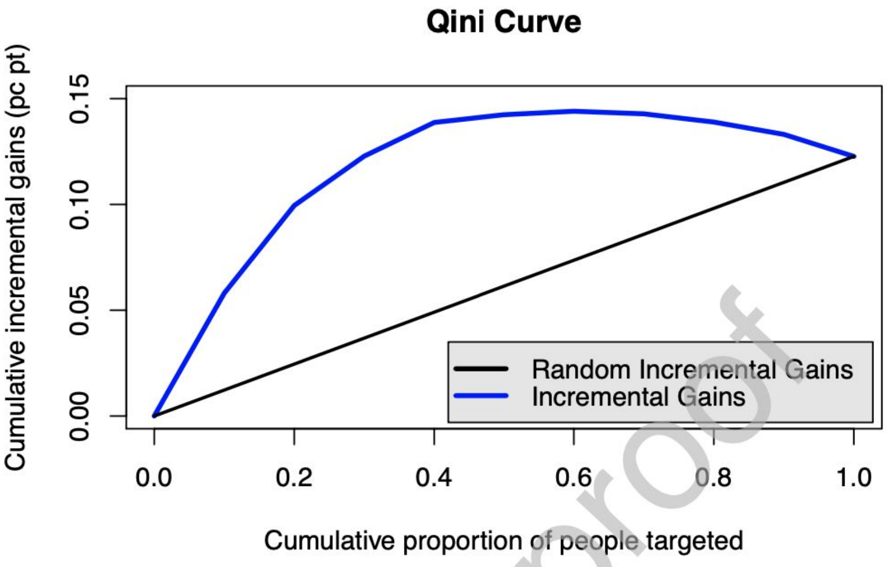
$$U(x_i) := \underbrace{P(y_i = 1 | x_i; t_i = 0)}_{\text{Probability of churn without treatment}} - \underbrace{P(y_i = 1 | x_i; t_i = 1)}_{\text{Probability of churn with treatment}}$$

The method proposed here would group all customers into a single dataset to be trained on. This would include the dummy variable indicating if a customer was in a treatment group or not, interactive variables between predictor and dummy variables (such as percent discount offered in treatment) and the predictor variable, indicating if a customer churned (1).

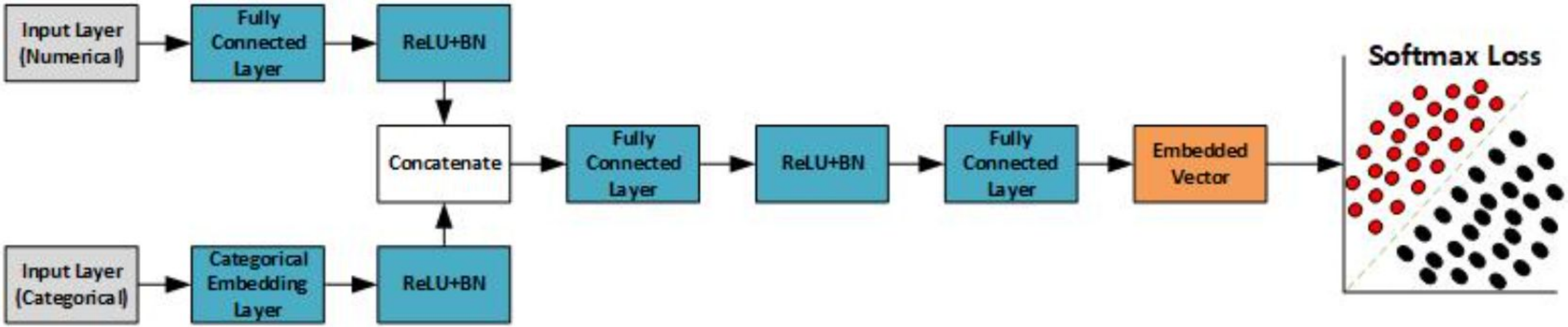
Evaluation of Uplift model:
To evaluate uplift, we must assess the ability of the model to identify groups of customers for whom treatment results in a substantial net effect.

The Qini curve helps to visualize the effect of outreach to increasing portions of the customers ranked by the uplift model from high to low net change in churn probability given treatment **(5)**.

The decrease seen in the plot below indicates capturing of lost causes, decreasing overall gains.

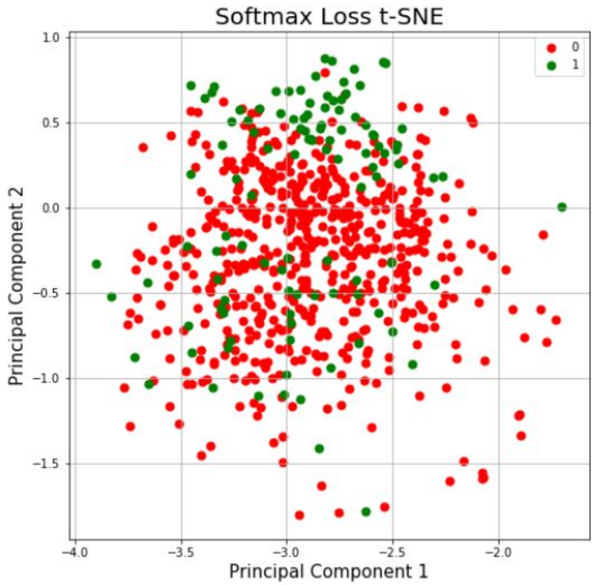


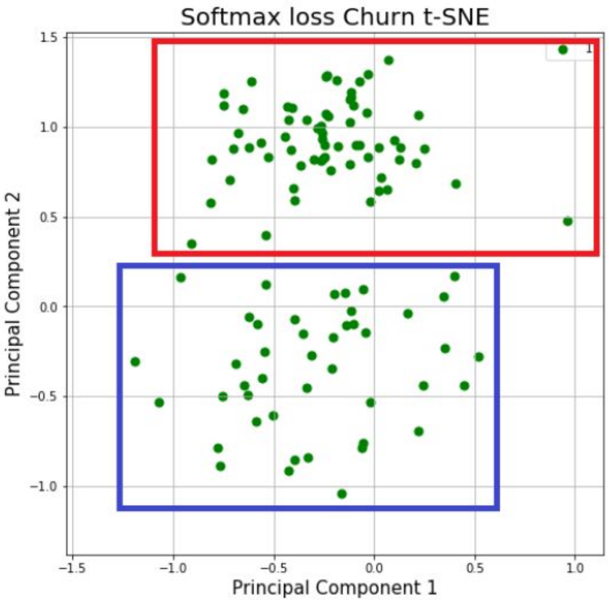
Deep Learning Model to Extract Vector Embeddings:
The deep learning model would be a series of fully connected layers with a vector embedding generating layer at the end. The final embedded vector can be analyzed with a t-SNE plot (dimension reduction technique) to understand churn behaviors relative to loyal customers.



Deep learning to predict churn is currently used for customers in the retail industry, music streaming service, mobile gaming and the telecommunication industry **(2)**.

Evaluating Vector Embeddings:
Previous work done using deep learning techniques in the telecommunications industry was able to produce the following plots. They reveal the behavior of churned customers forming two groups, one separated from loyal customers and one with similar behavior to loyal customers.



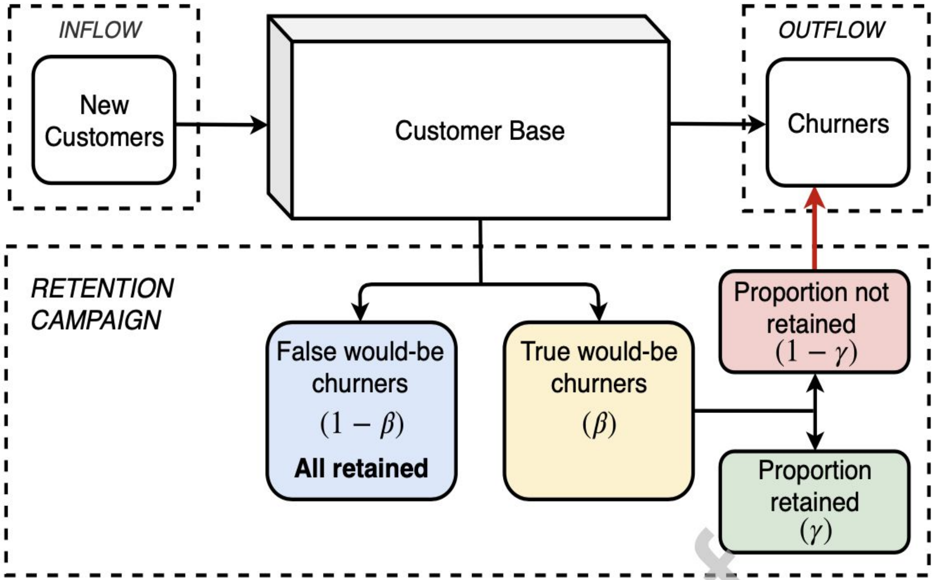


The plots can be evaluated using intra and inter cluster cosine similarity to determine compactness and relative distance between clusters **(2)**.

Comparison point between Deep Learning Model and Uplift Model:
An adapted version of the maximum potential profit can be used to compare models.

The maximum profit formula calculates the ROMI (Return on Market Investment) for a given set of predictions. The varying parameter is the portion of customers targeted (α), the remaining terms represent costs of outreach, penalizing for customers who churned regardless (lost causes) and ones that would not have churned but took advantage of the offer anyways.

$$MP = max_{\alpha}(\prod)$$



$$\Pi = N\alpha[\ \underbrace{\underbrace{\beta\gamma\,(b - c_{\text{contact}} - c_{\text{incentive}})}_{\text{Retained churners}} + \underbrace{\beta(1-\gamma)\ (-c_{\text{contact}})}_{\text{Non-retained churners}}}_{\text{True would-be churners}} + \underbrace{(1-\beta)\,(-c_{\text{contact}} - c_{\text{incentive}})}_{\text{False would-be churners}}\ ] - A$$

Data Needs:
Variables not currently provided by dockers that would serve to improve model performance

- **Customer transaction history**
- **Outreach campaign discount for targeted groups**
- **Data on outcomes of a prior treatment campaign**

Customer transaction history was found to be a significant indicator of churn in the retail industry while using deep learning models **(3)**.

To improve manipulation capabilities for the Uplift model, the discount offered in the target campaign is needed. It can be used as a factor in the treatment group to determine thresholds for maximizing the number of persuadable customers.

References:

1. “Why you should stop predicting customer churn and start using uplift models” https://www.sciencedirect.com/science/article/pii/S0020025519312022
2. “Deep Learning as a Vector Embedding Model for Customer Churn” https://www.sciencedirect.com/science/article/pii/S1877050921000557
3. “Comparison of Deep Learning Algorithms to Predict Customer Churn within a Local Retail Industry” http://www.ijmlc.org/vol7/634-P16.pdf
4. “Defection detection: Measuring and understanding the predictive accuracy of customer churn models” http://wak2.web.rice.edu/bio/My%20Reprints/Defection%20detection.pdf
5. Using control groups to target on predicted lift: Building and assessing uplift model. / Radcliffe, Nicholas. In: Direct Marketing Analytics Journal, 2007, p. 14-21.
6. W. Verbeke, D. Martens, C. Mues, B. Baesens, Building comprehensible customer churn prediction models with advanced rule induction techniques, Expert Syst. Appl. 38 (2011) 2354–2364.
7. M. R. Colgate, P. J. Danaher, Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution, J Acad Market Sci 28 (2000) 375–387.
8. J. Ganesh, M. J. Arnold, K. E. Reynolds, Understanding the customer base of service providers: An examination of the differences between switchers and stayers, J Mark 64 (2000) 65–87.