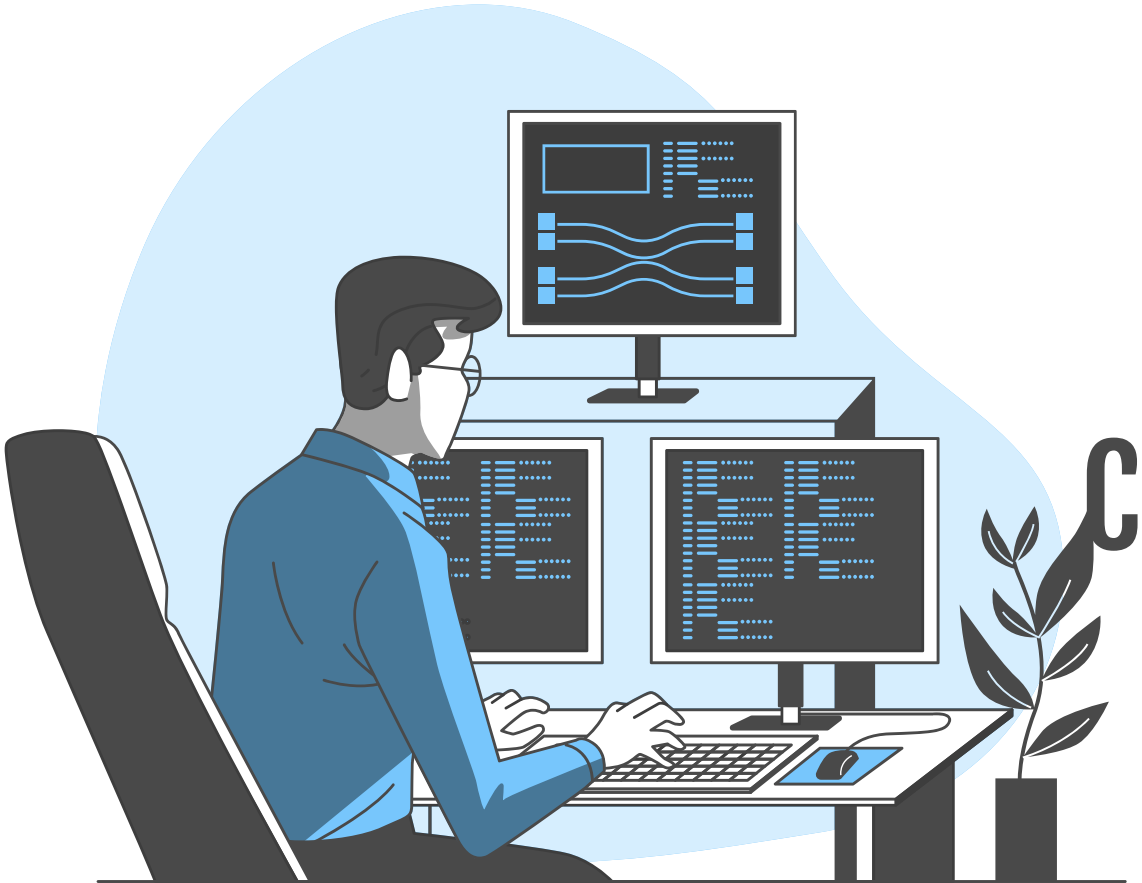
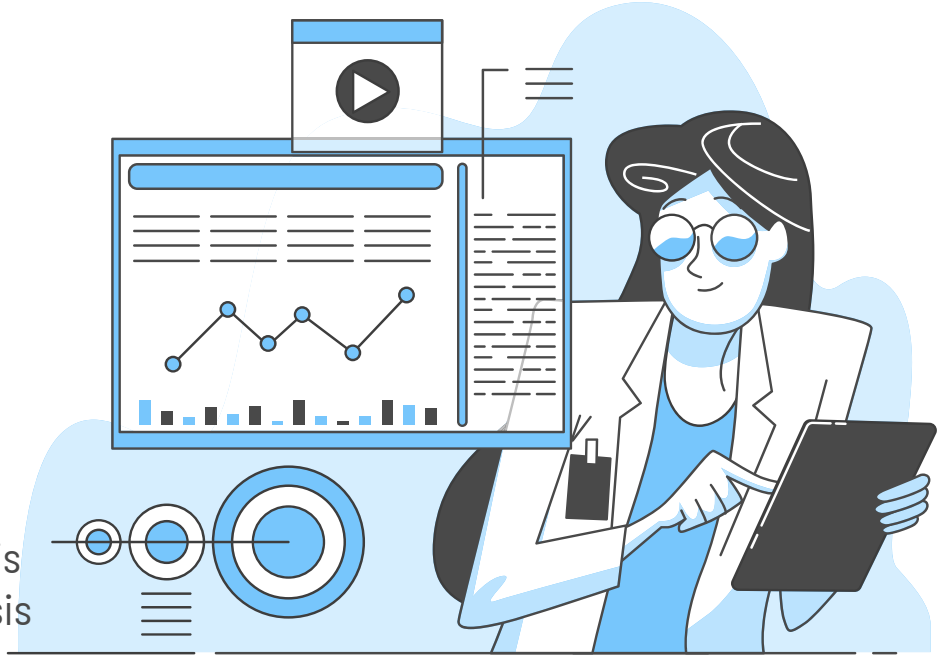


Kaggle Competition

MSDS 630

Karsten & Nestor

Table of Contents



Feature Created

01 Unseen User Embedding

Simulated User for users not seen during training by matching items rated

02 Negative Sampling

Created User and Item pairings and set rating to 0.

03 Concatenation

Neural Network Embeddings of Users and Items were concatenated.

04 Context Feature ID

We did not utilize this in our best model.

Techniques Applied

Matrix Factorization

Learning Rate: 0.001
Embedding Size: 100
Epochs: 25

Random Forest

Trees: 300
Max Depth: 8
Min Leaf: 10

Neural Network

Learning Rate: 0.001
Embedding Size: 10
Epochs: 20

Drop Out

20%

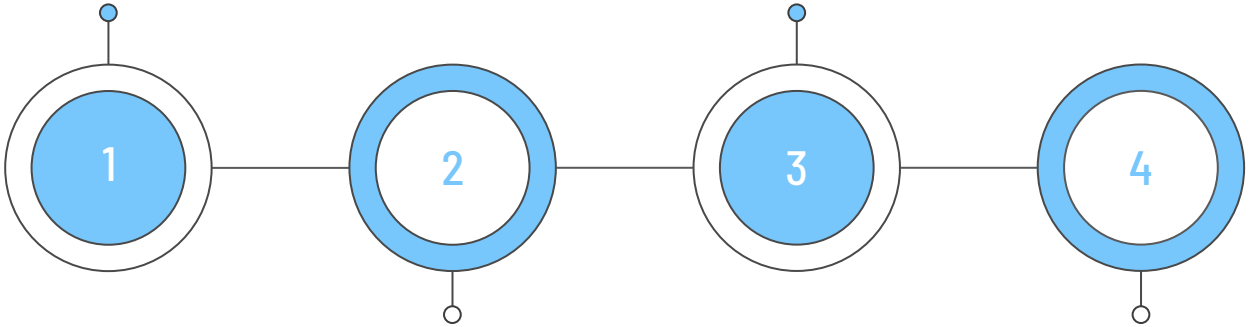
Negative Sampling

Ratio: 6-to-1

Our Experiments

Matrix Factorization
Hyperparameter Tuning

Implement Random Forest
Fine Tune Negative Sampling
Hyperparameter Tuning

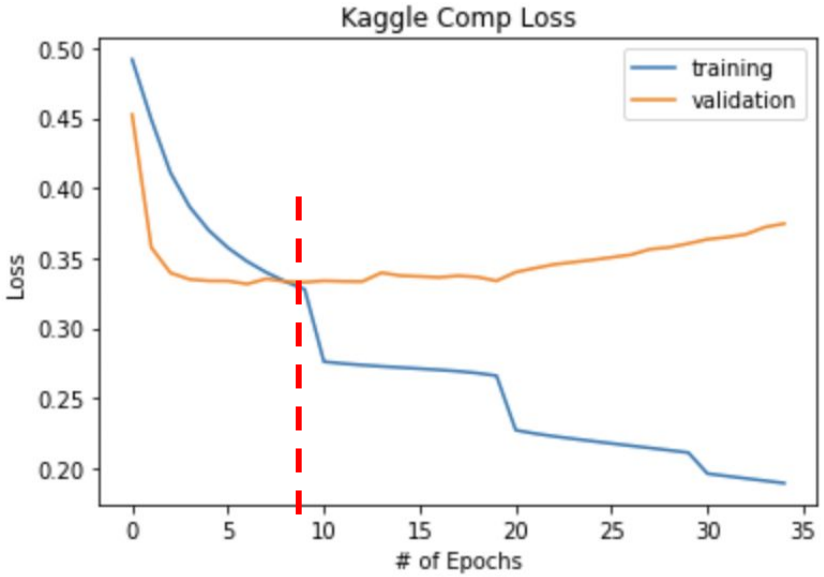


Introduce Bias
Introduce Negative Sampling

Neural Network
Introduce Dropout

Results

	Matrix Factorization with bias	Random Forest	Simple NN	Deep NN w/ drop
Training Loss	.436		0.182	0.324
Validation Loss	.470		0.315	0.333
Validation Acc.	.812		0.813	0.799
Leaderboard Score	.41293	.484	.468	.448



Severe overfitting occurring as
the number of epochs increases.

Leaderboard

DudeWheresMyModel

1st: 0.40290

Overfit Bros

2nd: 0.40607

NesKar Go Vroom!!

13th: 0.41293