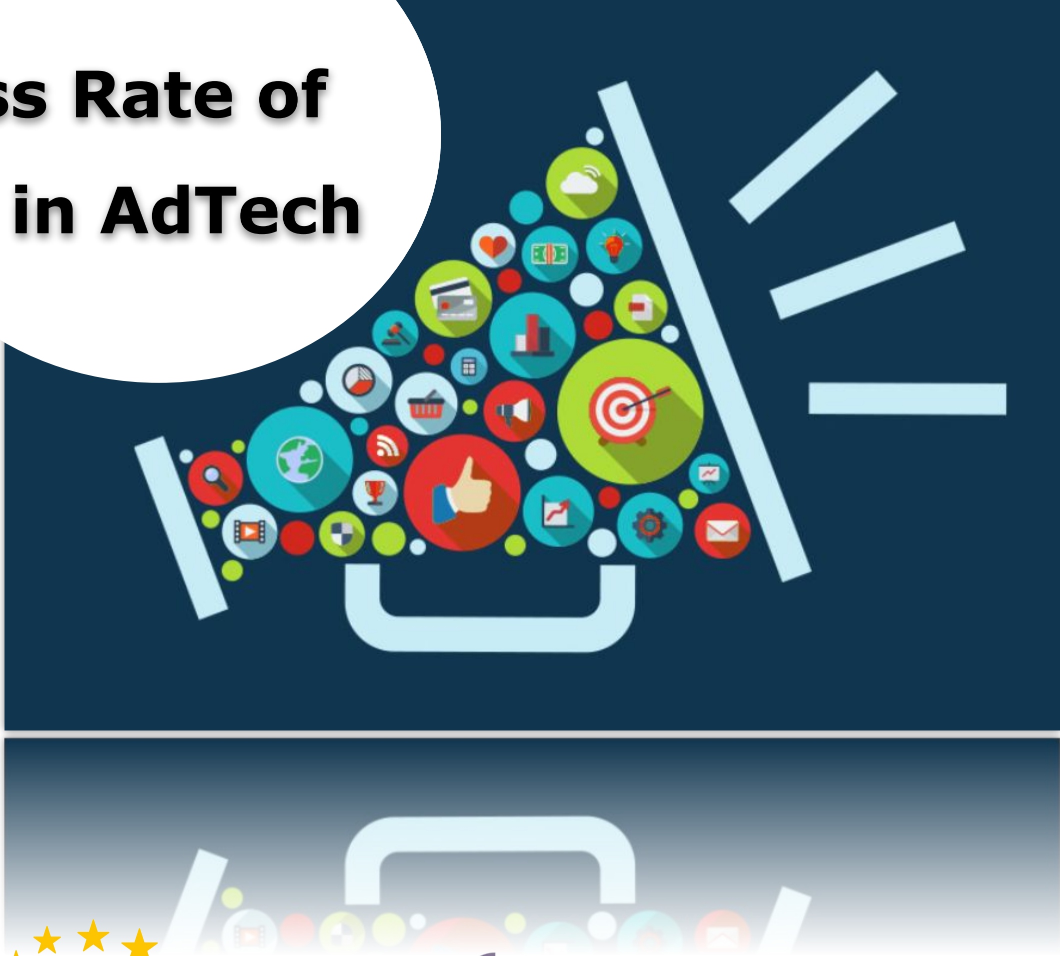
Predicting the Success Rate of Video Advertisements in AdTech



Buse Ozer

September 2020

Advisor: Boris Prodhomme

Tutor: Nacera Bennacer

BDMA

CentraleSupélec

smart

Agenda

1. Introduction to Digital Advertising
2. General Overview of Advertising Ecosystem
3. Problem Definition
4. Methodology
5. Comparison of Different Models
6. Conclusion
7. Future Work

CentraleSupélec

smart

Introduction to Digital Advertising

Why Digital Display Advertising Is Important ?

High increase in digitalization

High revenue in digital advertising

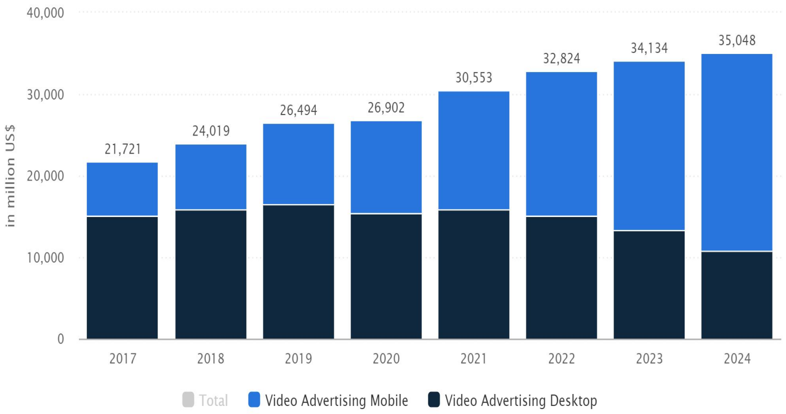
Helps to accelerate digital commitment of users

Why Are Video Advertisements Important ?

Drives more engagement

Leads to more revenue

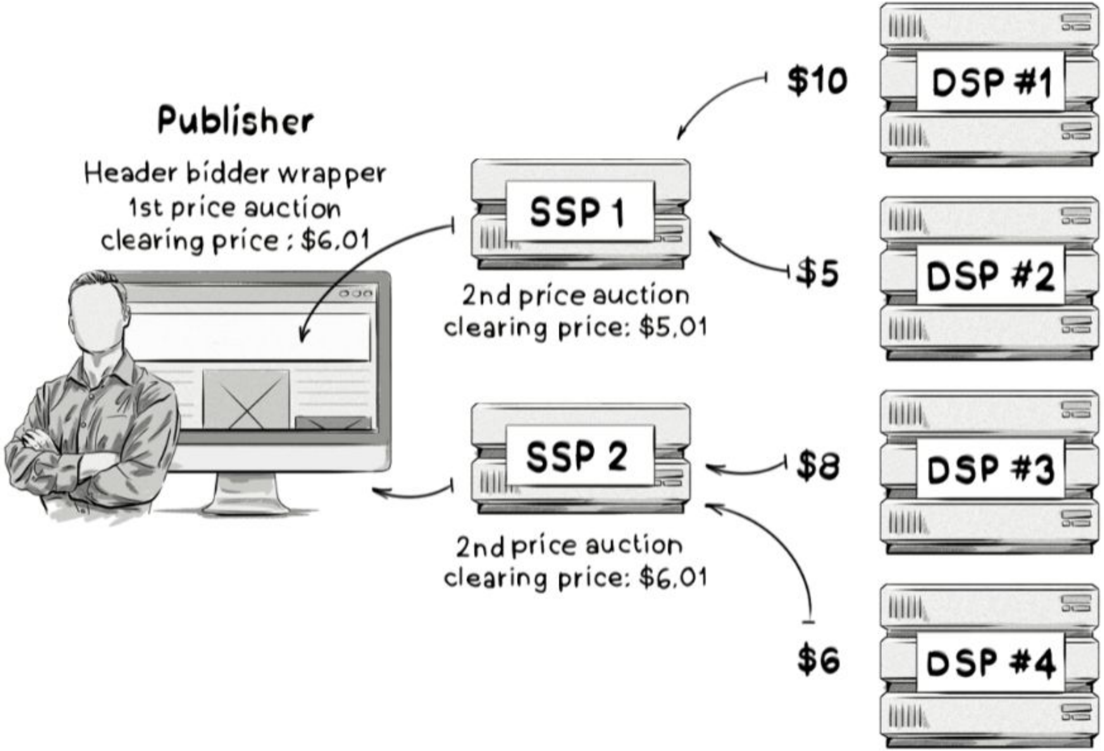
Better user experience



CentraleSupélec

smart

General Overview of Advertising Ecosystem



Problem Definition

Success Rate Definition

Probability of an ad to be delivered

$$\text{Success rate} = \frac{\text{\# of times a video ad displayed}}{\text{\# of times a video ad won an auction}}$$

Why Do We Focus on Video Success Rate ?

Our Display RTB delivery rate > 90 %

Our Video RTB delivery rate < 30%

Why Success Rate Optimization Is Important for Smart ?

Each loss opportunity is a direct loss in revenue for us and our publishers

CentraleSupélec

smart

The Video Obstacles

Problems

- ❏ Timeouts of media files
- ❏ Player and creative incompatibility
- ❏ VPAID issues
- ❏ User leaves the page before the ad loads
- ❏ Connection cut

Solutions

- ❏ Ensure that video ads run smoothly
- ❏ Predict the delivery rate of a bid
- ❏ Highest expected revenue

Bidder	Bid price	Delivery rate (predicted)	Expected revenue
A	$10	10%	?
B	$4	50%	?

The Objective

Increase of Smart and Publisher's revenue!

1. Apply delivery rate optimization
2. Increase the delivery rate of publishers
3. Increase the revenue of publishers
4. Increase the revenue of Smart

CentraleSupélec

smart

Predictive Model

What exactly is to be predicted ?

The probability of an ad to be displayed to an end-user

What is the goal of the predictor ?

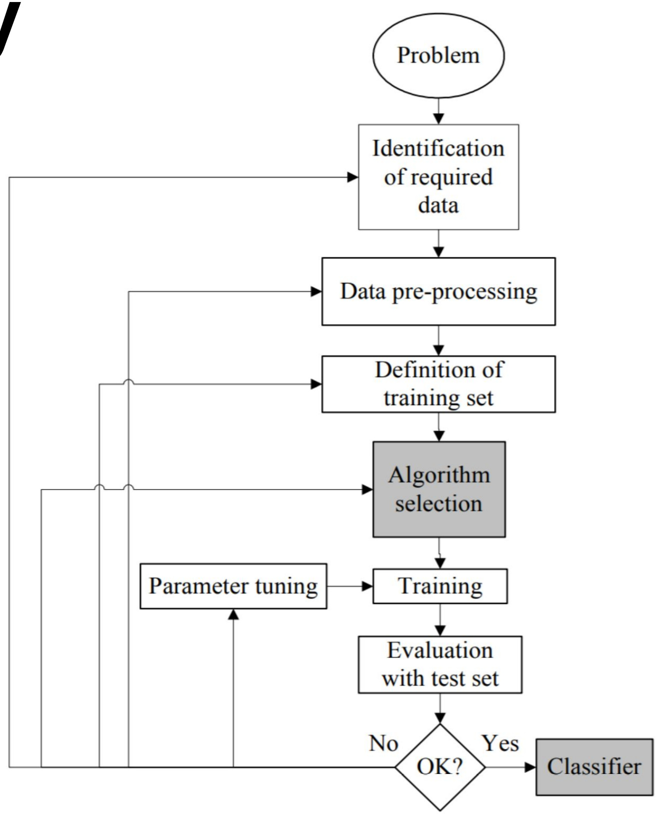
1. Be as **accurate** as possible
2. Have as **a few classification errors** as possible

What is the metric to measure the model performance ?

AUC, ability of a classifier to distinguish between classes

Among all “positive”-“negative” pairs in the dataset compute the proportion of those which are ranked correctly by the evaluated classification algorithm.

Methodology



CentraleSupélec

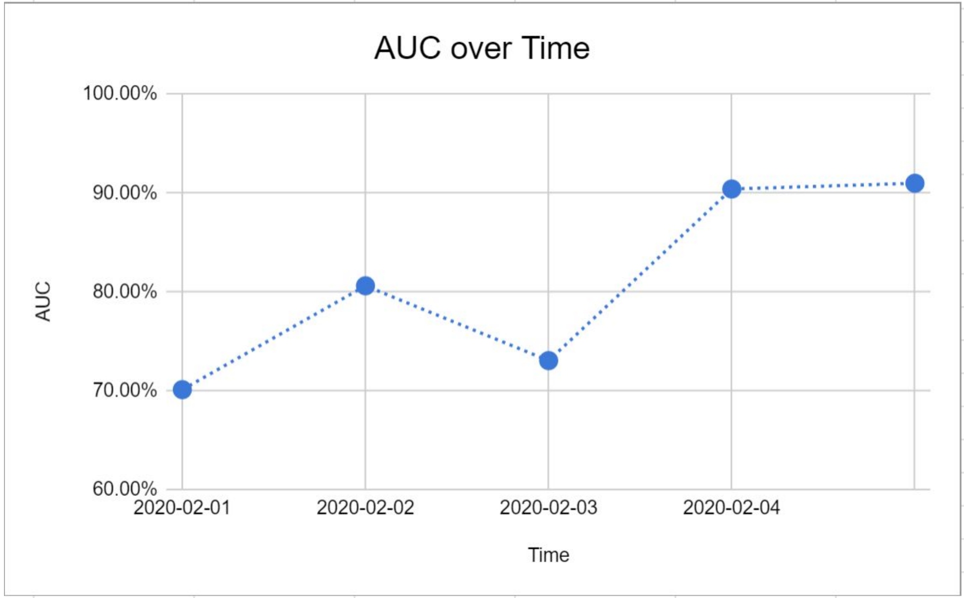
smart

Data Set Split

DATA

Training | **Dev** | **Test**

Experiment	Training Set	Dev Set	Test Set
Experiment 1	Day 1	Day 2	Day 3
Experiment 2	Day 2	Day 3	Day 4
Experiment 3	Day 3	Day 4	Day 5
...			
Experiment n	Day n	Day n+1	Day n+2



Comparison of Tree-Based Models

Data Set	Model	LogLoss	Precision	Recall	Accuracy	AUC	F1
2020-02-01	Decision Tree	0.7477	0.8820	0.6223	0.6726	0.7942	0.7297
2020-02-01	Ada Boost	0.7259	0.8798	0.6099	0.6634	0.7705	0.7204
2020-02-01	Random Forest	0.6130	0.8789	0.6176	0.6678	0.7907	0.7254
2020-02-01	XGBClassifier	0.5960	0.8869	0.6242	0.6766	0.8059	0.7327
2020-02-02	Decision Tree	2.4462	0.7590	0.9019	0.7287	0.6389	0.8243
2020-02-02	Ada Boost	0.7950	0.7621	0.8882	0.7247	0.6562	0.8203
2020-02-02	Random Forest	0.4752	0.7496	0.9541	0.7452	0.7186	0.8396
2020-02-02	XGBClassifier	0.4716	0.7517	0.9358	0.7381	0.7330	0.8337
2020-02-03	Decision Tree	0.4975	0.8563	0.8996	0.8241	0.8728	0.8774
2020-02-03	Ada Boost	0.4750	0.8566	0.8960	0.8221	0.8431	0.8759
2020-02-03	Random Forest	0.3438	0.8548	0.9084	0.8285	0.8741	0.8808
2020-02-03	XGBClassifier	0.3081	0.8586	0.9062	0.8306	0.9042	0.8818

Comparison of Tree-Based Models

Model	Log Loss	AUC
Decision Tree	1.2305	0.7686
Ada Boost	0.6653	0.7566
Random Forest	0.4773	0.7945
XGBoost	0.4586	0.8144

Importance of Freshness of The Model

Objective

Impact of re-calculating the model as frequent as possible

Before

Trained on Day n-1

Tested on Day n

82% of AUC

Result

Hourly basis gives **4%** better AUC

The most current data

↓

Model performs better

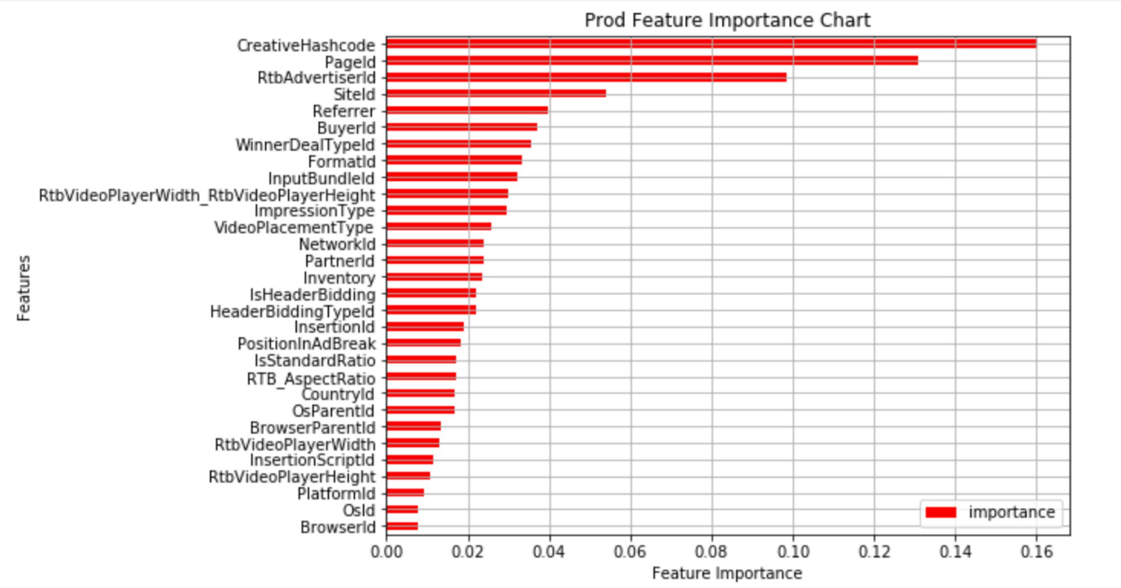
After

Trained on hour n-25, n-23, ..., n-1

Tested on hour n, n+1, n+2, n+3

86% of AUC

Adding New Dimensions



3.21 %

New Features from RTB

PageId	OsParentId
BuyerId	OsId
WinnerDealTypeId	BrowserId
FormatId	RtbVideoPlayerWidth
CountryId	RtbVideoPlayerHeight
InsertionId	VideoPlacementType
BrowserParentId	PositionInAdBreak

CentraleSupélec

smart

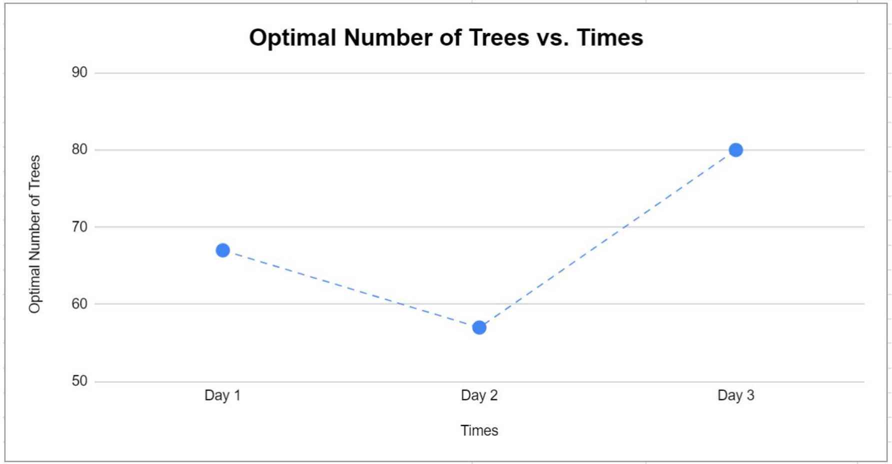
Early Stopping Strategy

Problem

There is a high variance between model complexity over time

Solution

Applying early stopping algorithm in order to control the model complexity



Early Stopping

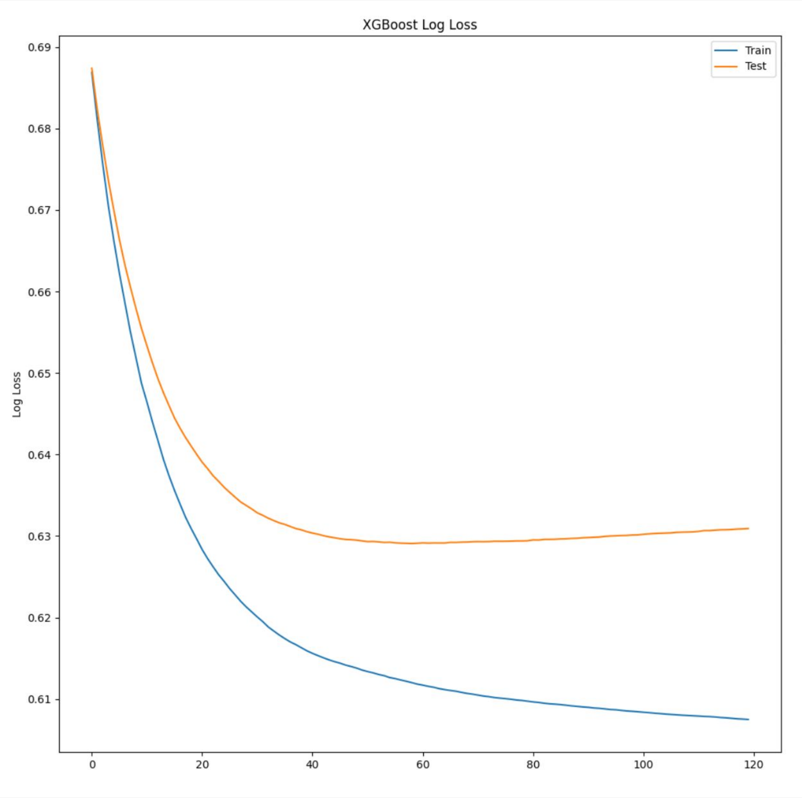
Use training and dev sets

Keeps on training the model until there is no more improvement on dev set

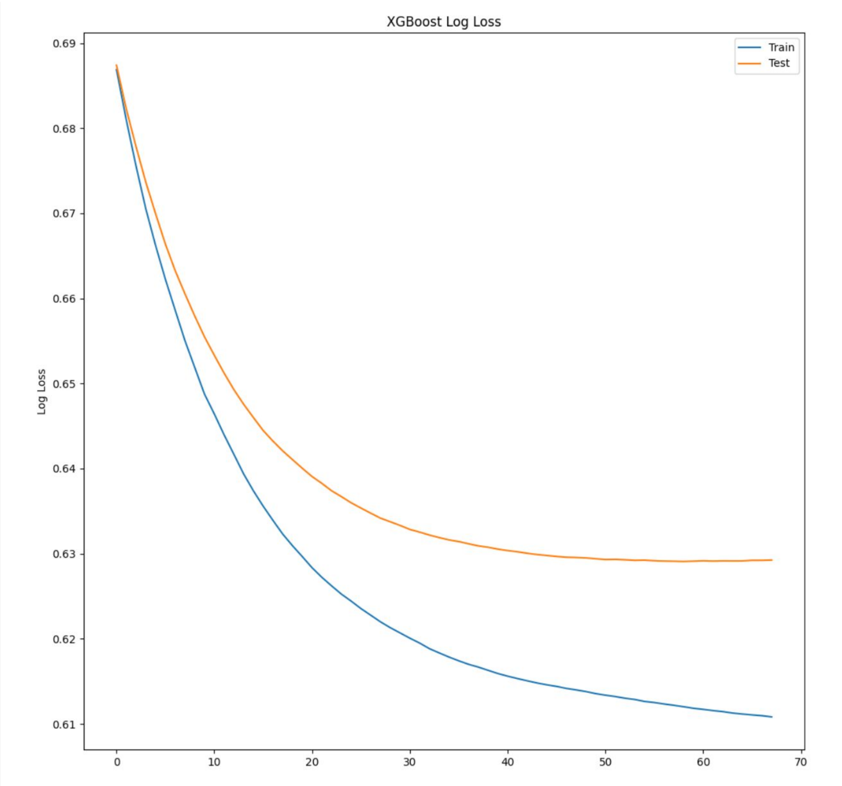
Determine how many trees to be built before overfitting

Prevents Overfitting

Without Early Stopping Algorithm



With Early Stopping Algorithm



Neural Network Model

Why Neural Network Model?

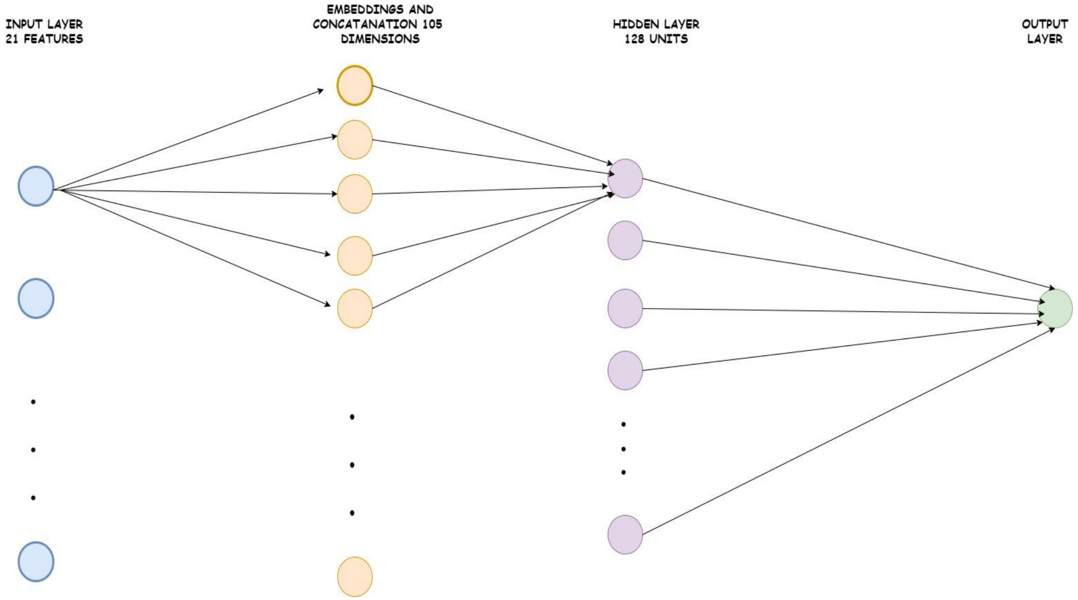
- Entity Embedding
- Ability to apply continuous learning
  - Ability of a model to learn continuously
  - Progressive and adaptive learning as new data comes in
  - Overcome forgetting of previously seen data
- Transfer Learning
  - The ability to transfer knowledge from one domain to another
  - Train a model on different tasks at the same time
- TensorFlow Serving in production

CentraleSupélec

smart

Neural Network Model

- Embedding layer with 5 dim
- 1 hidden layer with 128 units
- Dropout with 0.3 probability



TensorFlow

Keras

CentraleSupélec

smart

Final Results

XGBoost Model

Training AUC **92.4%**

Test AUC **89.8%**

Neural Network Model

Training AUC **91.5%**

Test AUC **89.2%**

CentraleSupélec

smart

Conclusion

- There is **no trend** in AUC over time
  - Run experiments on multiple days
- XGBoost model **outperforms** Decision Tree, AdaBoost and Random Forest
  - Better AUC, better **control over the model**
  - Scalable and parallelizable
- High variance in data
  - The model has to be **adaptive** and **flexible**
  - Early Stopping mechanism
- XGBoost and Neural Network performs nearly the same
  - Neural Network requires more expertise
- Implementation of Neural Network has some advantages
  - Continuous Learning
  - Serving in production
  - Transfer Learning

smart

Future Work

- Change the weighting strategy of the XGBoost model
  - Weighting on the revenue of samples to focus where the impact is high
- Continuous Learning in Neural Network Model
- Anomaly Detection
  - Detect spaces where delivery late is low
  - Finding the responsible publisher for the anomaly

Thank you !

Predicting the Success Rate of Video Advertisements

Buse Ozer

September 2020

CentraleSupélec

References

- Page 3, https://www.statista.com/outlook/218/100/video-advertising/worldwide
- Page 4, https://clearcode.cc/blog/sequential-auctions-header-bidding-first-price-second-price-auctions/
- Page 9, Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT), 48(3), 128-138.


Q&A