MULTIMODALITY AND STACKING ENSEMBLE MODELS IN DEMAND PREDICTION

Steve Zhou, Ellie Yang

MIT MANAGEMENT
SLOAN SCHOOL

PROJECT OVERVIEW

Problem Statement

Challenge:

- to predict deal probability for listings on Avito.com, Russia's largest classified advertisement platform

Goal:

- to equip sellers with insights to optimize listings and help the platform to improve user experience

Data Summary

Scope

- 1.5M listings in March 2017
- Tabular, text, image data

Target Variable

- *deal_probability*

Challenge

- Tabular mostly categorical
- Russian text
- Large size (50GB+ images)

Methodology

Multi-modality

- Text embeddings with TF-IDF, SpaCy, FastText and images embeddings with ResNet 50
- SVD dimensionality reduction

Stacking Ensemble

- Combine outputs from base models across data modalities

(TABULAR) FEATURE ENGINEERING

Original Features	Price Statistical Features	Text-Related Features	Combination Features
region	category_price_mean	title_length	region_city
city	category_price_std	description_length	all_category
parent_category_name	category_price_skew	title_word_count	category_param_1
category_name	city_price_mean	description_word_count	region_category_user
param_1	city_price_max	title_has_keyword	city_category_user
param_2	city_price_skew	description_has_keyword	
param_3	price_to_category_mean	title_digit_count	
price	price_to_category_max	description_digit_count	
user_type	price_log	description_newline_count	
	price_bin	description_missing	
TOTAL: 9	TOTAL: 10	TOTAL: 10	TOTAL: 5
TOTAL: 34			

Table 1: Tabular Feature Engineering

- **9** Original Features
- **10** Price Statistical Features
- **10** Text-related Features (length, word count...)
- **5** Combination Features

34

Total Tabular Features into future models

TEXT FEATURE EXTRACTION

- Extract Embeddings with three different methods
- Use Singular Value Decomposition (SVD) to reduce dimension

TF-IDF

- TF-IDF, 500 Embeddings
- TF-IDF, Embeddings SVD to 5d

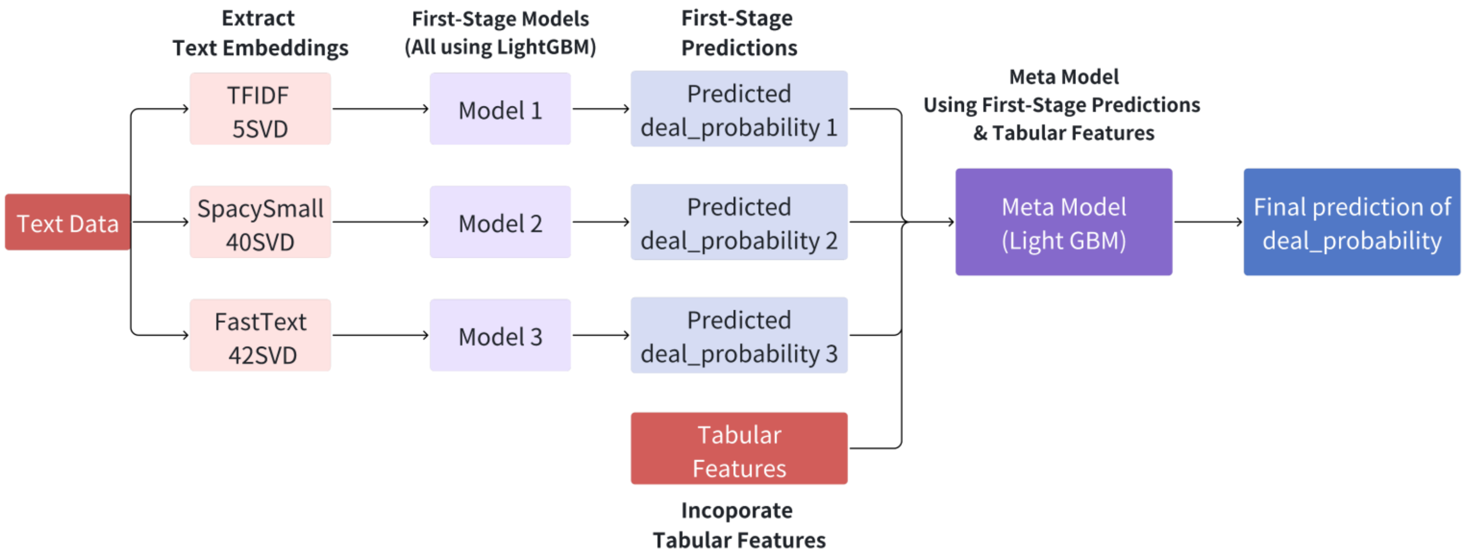
SpaCy

- SpaCy Small Russian, 98 Embeddings / Medium Russian, 300 Embeddings
- SpaCy Small Russian, SVD to 40d / Medium Russian, SVD to 100d

FastText

- FastText trained on our data, Embeddings 100
- FastText trained on our data, SVD to 42d
- Pretrained FastText, Embeddings 300
- Pretrained FastText, SVD to 50d

STACKING MODELS



Stacking Model 1:

- **3 First-Stage Models:**
  - TFI-DF, Spacy, FastText Embeddings→ Predict deal_probability separately
- **Meta Model:**
  - Combines first-stage predictions and tabular features

Stacking Model 2:

- **2 First-Stage Models:**
  - FastText → Predict price
  - Spacy → Predict deal_probability
- **Meta Model:**
  - Combines first-stage predictions, tabular features, and TFI-DF embeddings

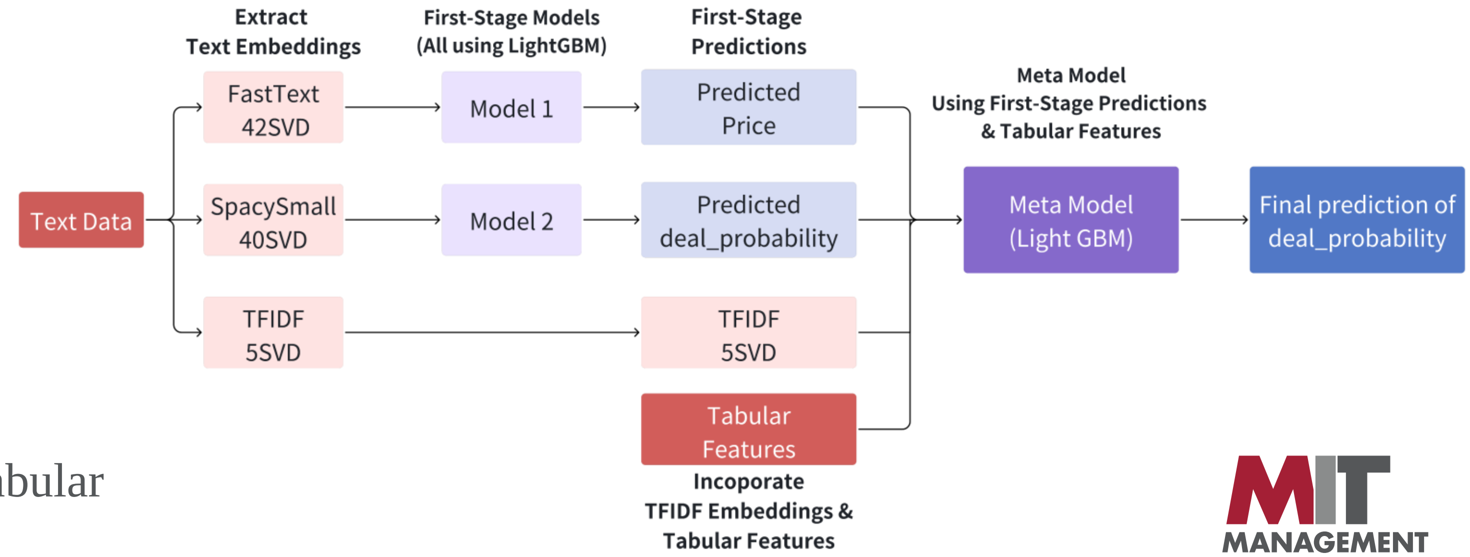


Table 4.1: Summary of Text and Tabular Models Performance in Complete Dataset

Tabular Baseline	.227894

RMSEs when Directly Using Embedding as Features for LightGBM

	SVD Dimension	RMSE - Embedding Only	RMSE - **With Tabular**
TF-IDF	5	.252066	.227283
SpaCy Small	40	.248780	.227110
SpaCy Medium	100	.246289	.227009
FastText	42	.235021	**.226067**
FastText Pretrained	50	.235337	.227278

RMSEs when Stacking Embedding LightGBM

First-stage Models	Second-stage Model	RMSE
deal_probability ~ SpaCy deal_probability ~ TF-IDF deal_probability ~ FastText	First-stage Preds + Tabular	**.226904**
price ~ FastText deal_probability ~ SpaCy	First-stage Preds + Tabular + TD-IDF	.227119

0.227894

RMSE of Tabular Baseline

6.7%

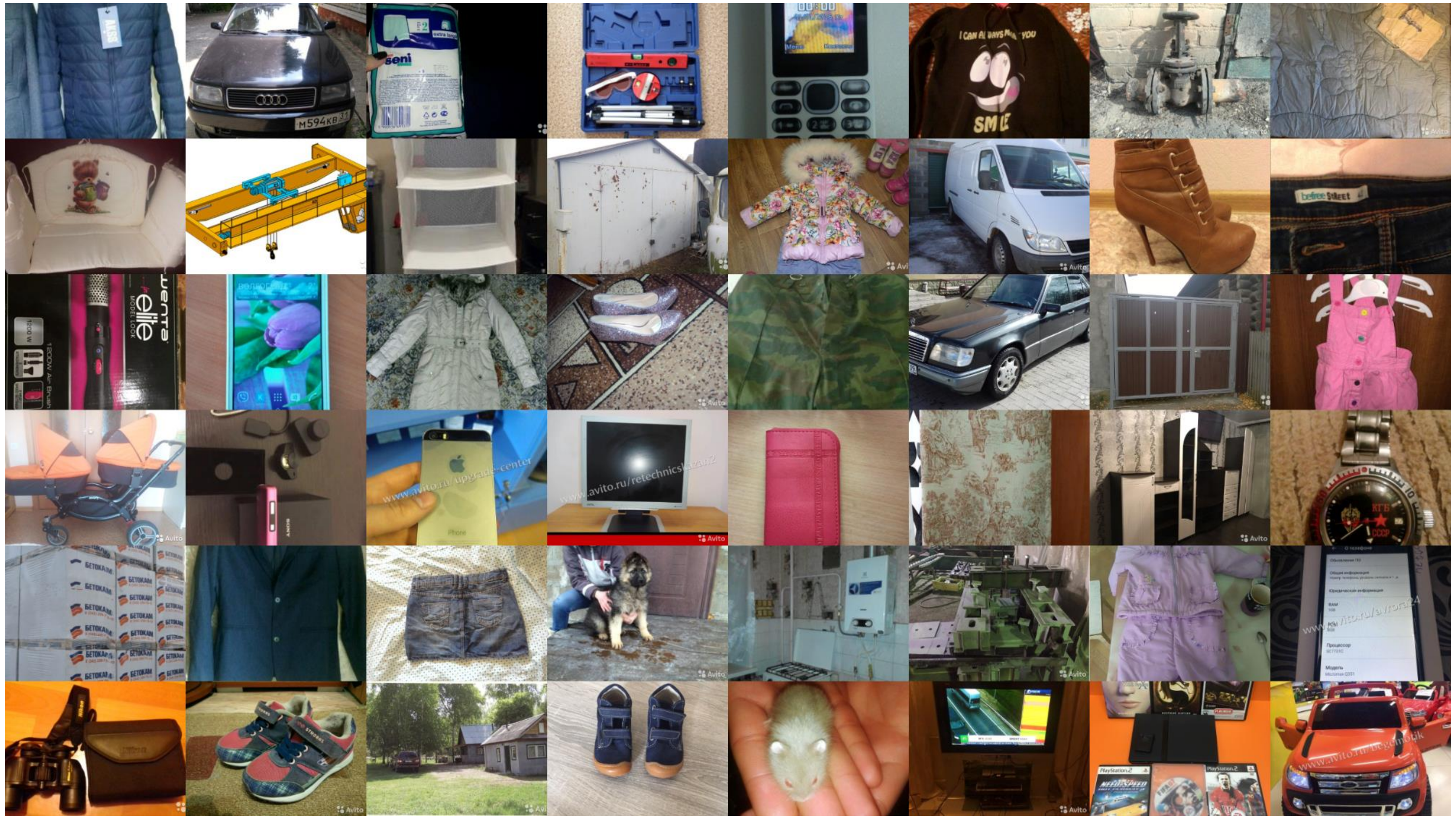
Average RMSE improvement By combining Tabular and Embeddings Features

0.226067

RMSE of Best Model

MANAGEMENT
SLOAN SCHOOL

IMAGE FEATURE EXTRACTION



- Visual features matter
  - Important product characteristics, product quality, and buyer interest
  - Majority (93%) of listings has associated image for the listed product
- Images are expensive
  - 50+ GB of image data, not realistic for our computing resources
  - 20% subset of data with images to evaluate image and other features

MIT MANAGEMENT SLOAN SCHOOL

MODEL PERFORMANCE IN IMAGE SUBSET

Tabular Baseline			.232013

RMSEs for Training and Testing Models in Subset for Image Data

	SVD Dimension	RMSE - Embedding Only	RMSE - **With Tabular**
Text Embedding	42	.241487	.230863
Image Embedding	500	.242552	.230489
Text + Image			**.229803**

.232013
RMSE of Tabular Baseline

.230489
RMSE of Best Bi-Modality

.229803
RMSE of Tri-Modality

CONCLUSION

- Our experiments demonstrate that
  - **Power of multimodality**: Inclusion of unstructured data is crucial for improving predicting *deal probability* of online listings
  - **Comparing text models**: Neural-network-based text embedding (FastText) outperforms frequency-based embedding (TF-IDF) in this task
  - **Complementary power**: Inclusion of image features on top of text and tabular models significantly boosted predictive accuracy
- Stacking ensemble
  - **Stacking boosted accuracy:** Effective in leveraging information from diverse sources

- Some constraints still require future work
  - **Feature Engineering**: Explore additional engineered features from tabular, text, and image data to enhance predictive performance
  - **Advanced Stacking**: Incorporate image embeddings alongside text embeddings in first-stage models, explore multi-stage stacking, or test new intermediate response variables
  - **Scaling Image Analysis**: Extend image feature analysis to the full dataset to fully capture the role of visual features in deal probability
  - **Model Extensions**: Experiment with more advanced architectures for image and text embeddings, such as transformers

THANK YOU

MIT MANAGEMENT
SLOAN SCHOOL