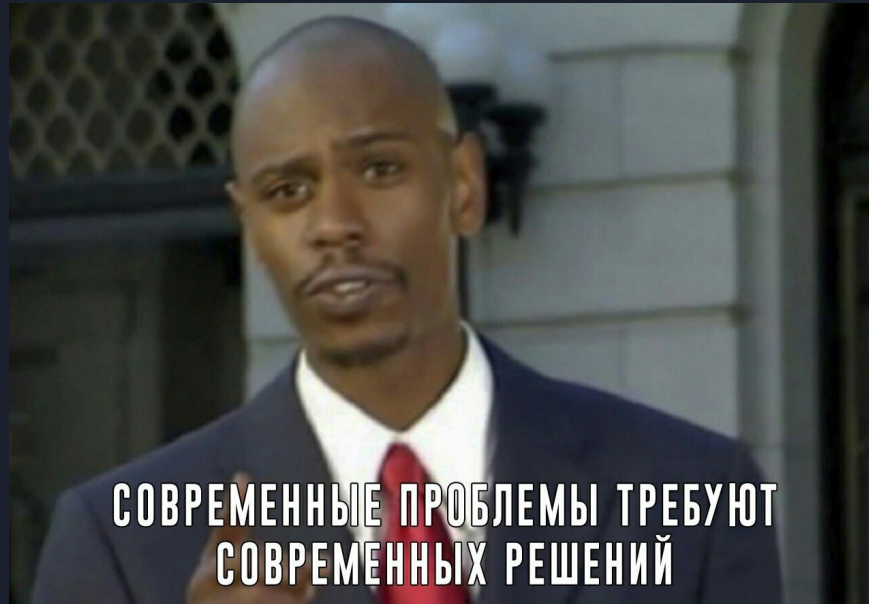




Анализ “стоимости игроков”

Проблема

Наша онтология решает проблему покупки игрока клубом



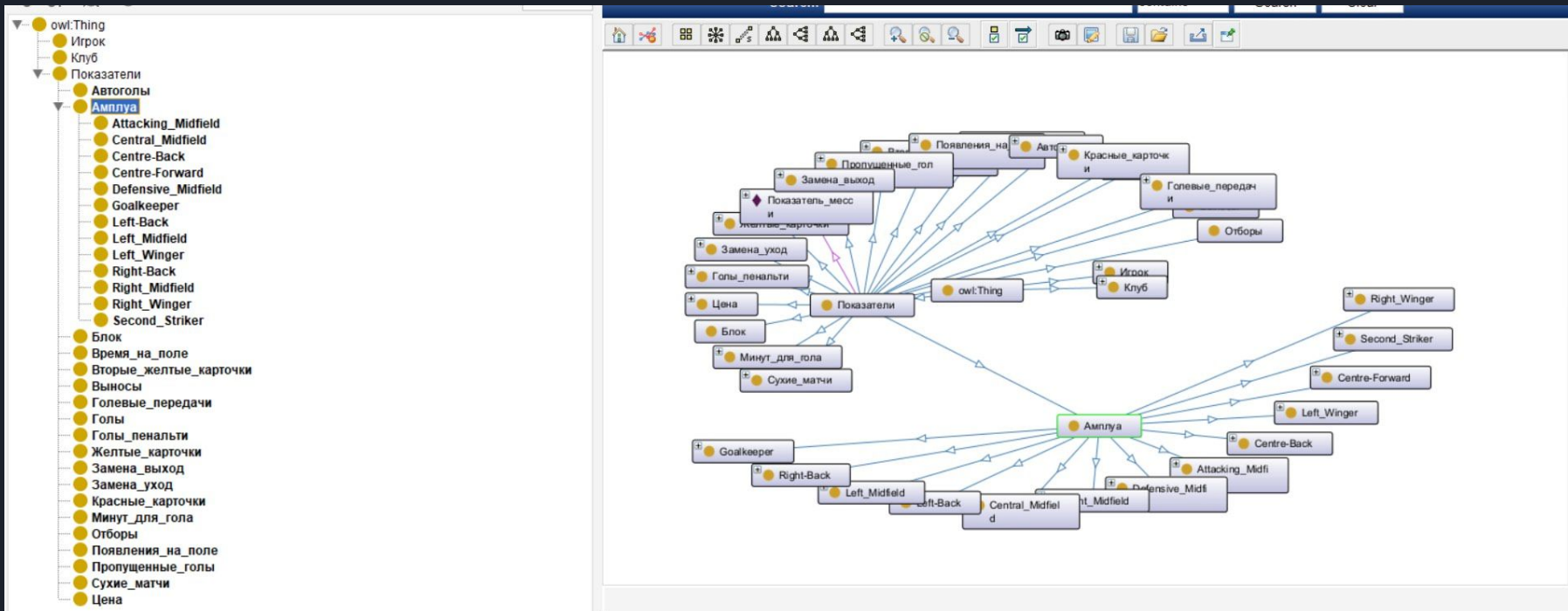


Исходные данные

Сайт: <https://www.transfermarkt.com/>

Парсер: https://github.com/andrey8080/Transfermarkt_parser

Онтология





Онтология

Количество сущностей: ~700

Количество классов: 35

Количество триплетов: ~9000



SPARQL

Каких нападающих и вратарей стоит купить если у клуба есть 1 000 000 евро?

Топ 10 нападающих

- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Ali_Jasim_player, mpg: 117, cost: 450000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Dailon_Rocha_Livramento_player, mpg: 141, cost: 800000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Mathis_Lamblond_player, mpg: 142, cost: 500000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Mario_Balotelli_player, mpg: 151, cost: 500000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Christian_Gytkjaer_player, mpg: 164, cost: 700000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Anas_Haj_Mohamed_player, mpg: 175, cost: 200000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Mateusz_Kowalski_player, mpg: 177, cost: 300000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Jeppe_Corfitzen_player, mpg: 191, cost: 400000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Leonardo_Pavoletti_player, mpg: 193, cost: 800000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Mirko_Maric_player, mpg: 202, cost: 800000

Топ 10 вратарей:

- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Pepe_Reina_player, clean_sheets: 359, cost: 600000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Leandro_Chichizola_player, clean_sheets: 126, cost: 400000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Alberto_Palmeri_player, clean_sheets: 120, cost: 500000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Nicola_Leali_player, clean_sheets: 105, cost: 600000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Matteo_Grandi_player, clean_sheets: 103, cost: 150000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Raffaele_Di_Gennaro_player, clean_sheets: 99, cost: 300000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Carlo_Pinsoglio_player, clean_sheets: 93, cost: 200000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Antonio_Donnarumma_player, clean_sheets: 89, cost: 325000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Nikita_Contini_player, clean_sheets: 81, cost: 400000
- Player : http://www.semanticweb.org/msi/ontologies/2024/10/untitled-ontology-12#Alen_Sherri_player, clean_sheets: 81, cost: 350000



Embedding

В результате обучения эмбедингов была решена проблема классификации по предсказанию диапазона цены игрока с точностью около 52.6% при прямом использовании параметров и one-hot-encoding в XGBoost классификаторе и около 39% при использовании эмбедингов игроков в качестве параметров. Точность второй модели заметно меньше модели, использующей детализированные признаки one-hot-encoding, что в нашем случае вероятнее всего связано с недостаточным объемом данных для обучения

Также была решена задача предсказания цены заданного игрока, демонстрирующая близкие к правдивым результаты

Наша модель на основе характеристик игрока может предсказать его стоимость



Вывод

Мы сделали классную работу при прохождении предмета “Графы знаний”, построили онтологию в protege, наполнили граф с помощью библиотеки RDFlib, составили SPARQL запросы, решили проблему классификации