Проблемные вопросы применения IT для решения задачи прогнозирования

Научный руководитель Мамай Дарья Сергеевна

Минск, 2018

Введение. Основные понятия и определения.

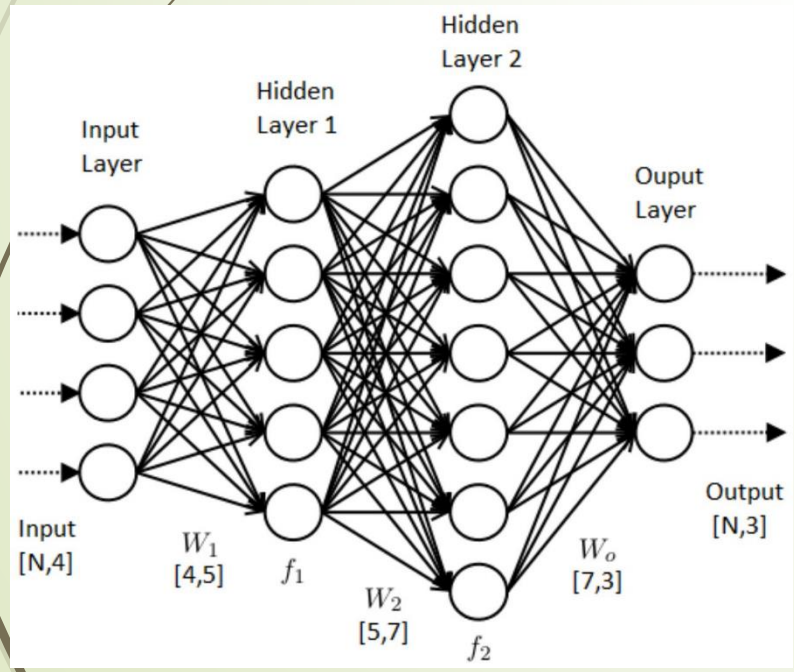
Машинное обучение (англ. machine learning, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач.

Дадим определение задачи теннисного прогнозирования.Пусть имеется следующая информация о предстоящем теннисном матче: имена участников, тип покрытия и позиции игроков в рейтинге ATP; а также историческая информация об уже сыгранных матчах. Необходимо построить модель на основе исторических данных, которая будет наилучшим образом прогнозировать результаты предстоящих матчей.

Обучающая выборка - набор данных для построения, обучения и анализа алгоритма машинного обучения. В случае задачи теннисного прогнозирования данная выборка представлена ранее сыгранными теннисными матчами и информацией о них.

Общая схема решения задачи прогнозирования

1. Поиск и фильтрация данных, построение обучающей выборки;
2. Отбор наиболее релевантных признаков для построения обучающей выборки;
3. Построение и обучение модели;
4. Анализ полученных результатов.



Построение обучающей выборки

Для построения обучающей выборки из источника 1 была загружена информация об мужских одиночных теннисных матчах, сыгранных под эгидой ATP, за 2004-2018 годы. Для фильтрации данных и построения выборки использовался фреймворк Pandas.
Pandas – библиотека языка Python, предназначенная для обработки и анализа данных. Работа библиотеки построена поверх библиотеки NumPy. Библиотекой предоставляются разнообразные структуры и алгоритмы числовыми данными и временными рядами.

Построение и обучение алгоритмов.

Алгоритмы для решения поставленной задачи:

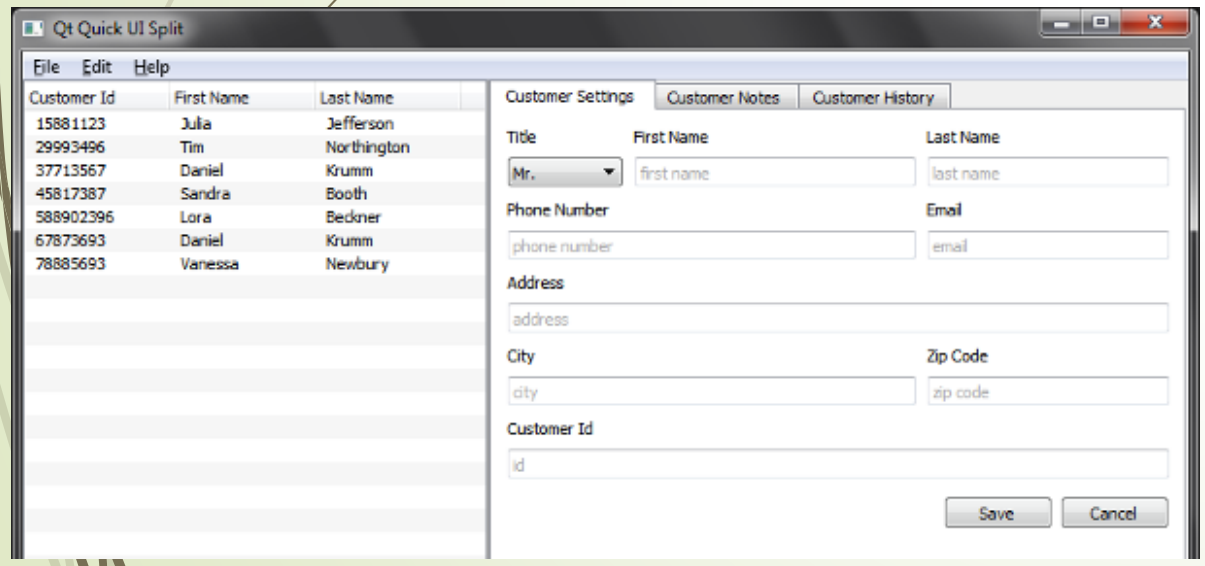
1. Логистическая регрессия;
2. Искусственная нейронная сеть (ИНС)
3. Метод опорных векторов

Keras – открытая библиотека на языке Python, предназначенная для построения нейросетевых моделей. Представляет собой интерфейс, надстроенный над Deeplearning4j, TensorFlow и Theano.

Scikit-learn – библиотека для машинного обучения, написанная на языке Python. В нее включены разнообразные алгоритмы классификации, регрессии и кластеризации, такие как случайный лес, метод опорных векторов, k-means, DBSCAN и другие.

Построение графического интерфейса

Qt — кроссплатформенный фреймворк для разработки программного обеспечения на языке программирования C++. Есть также «привязки» ко многим другим языкам программирования: Python — PyQt, PySide; Ruby — QtRuby; Java — Qt Jambi; PHP — PHP-Qt и другие.



Средства для написания реферата

Latex — наиболее популярный набор макрорасширений (или макропакет) системы компьютерной вёрстки TeX, который облегчает набор сложных документов. В типографском наборе системы TeX форматируется традиционно как Latex.