**Име на проекта**

**Резюме**

**Въведение**

Как да се предвиди цената на акциите, все още е горещ изследователски проблем за инвеститорите и изследователите във финансовата област. Прогнозирането на цените на акциите се превръща в изключително предизвикателна задача поради високия шум, нелинейността и нестабилността на данните от времевите редове на цената на акциите. Основната идея на всички усилия в тази насока е хипотезата за ефективния пазар, която казва, че цената на акциите може да бъде предвидена от историческите данни за търговията. Специалистите все по-често поглеждат към широкия набор съвременни възможности предоставяни от дълбокото машинно обучение и все по-успешно прилагат методите му. Повечето модели ползват разточителен набор от данни, включващи данни за търговията на конкретната акция, данни за индекси, валути, финансово-счетоводна информация, данни за макроикономически показатели и.н. Част от моделите се обучават и върху данни свързани със сентимента на пазара, като новини, туитове, постове в социални мрежи. Има конструирани и сентимент индикатори, които могат да бъдат ползвани директно в съвременните невронни мрежи.

В този проект е изграден модел за прогнозиране на времеви редове, наречен LSTM-GRU, който e хибридна мрежа състояща се от невронна мрежа за дългосрочна и краткосрочна памет (LSTM) и затворена повтаряща се невронна мрежа (GRU) за прогнозиране на цената на затваряне за определена ценна книга.

**Методология**

1. **Данни**

Модела бива захранван с два вида данни. През LSTM слоевете се прекарват данните (под формата на времеви редове с 30 дневни прозорци на цената на затваряне), както и седем изчислени индикатори. Във вторият сет са събрани данни за 85 други характеристики, включващи индекси, валути, инфлационни индикатори, криптовалути, съкровищни бондове, ETFs, взаимни фондове, както и други 50 компании от S&P 500. Пълен списък със ползваните данни, както и източниците на данни са предоставени в get\_data\data\indicators\_list.csv. Основния източник на данни е Yahoo Finance. Данните са подложени на задълбочен анализ, като допълнително е изготвен анализ за корелации, хетероскедастичност, серийни корелации и др. Направена е оценка на важните характеристики посредством PCA и XGBoosting. Подбрана е извадка от 2267 наблюдения обхващащи периода от 2013 до 2021 г. Данните са разделени на трениращи и тестови в съотношение 85%:15% (1926, 341). Част от трениращите данни (около 20% - 22%) се ползват за валидация.

Основни моменти от събирането, анализа, визуализацията и предподготовката на данните може да се намерят в get\_data\SMP\_GateringData.ipynb и get\_data\ SMP\_Data\_Analysis.ipynb.

Част от предподготовката на данните е изнесена в основният файл (SMP\_Price\_Predictor.ipynb в главната директория на проекта). Фондовите пазари са характерни с оформянето на дългосрочни трендове. За да се избегне заучаването на такива зависимости от модела, данните за трениране са умишлено размесени. Тази техника спомага да се спестят други методи, като cross validation, за обучение на модели с малки набори от данни. Данните за тестване следват времевата си хронология, за да може модела да бъде оценен максимално близо до евентуални реални работни условия. Следва min max нормализация на данните и преоразмеряване, подходящо за вход в невронните архитектури.

1. **Архитектури**

**Experiments**

1. **Метрики**
2. **Резултати**
3. **Анализи**

**Заключение**

**Бъдещи идеи за разработване**

**Ползвани източници**