**СОФИЙСКИ УНИВЕРСИТЕТ  
„СВ. КЛИМЕНТ ОХРИДСКИ“**

****

**ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА**

**КУРСОВ** **ПРОЕКТ**

**ПО СИСТЕМИ, ОСНОВАНИ НА ЗНАНИЯ**

**Тема:**

**"Модел за прогнозиране на бъдещото потребление на електроенергия въз основа на исторически данни".**

Изготвен от:

Стоян Тинчев Тинчев 3MI0700193

Кристиан Йорданов Васиков 0MI0700194

Специалност „Информационни системи“, курс 3; група 2

София, януари 2025 г.

Съдържание

[1. Въведение в проекта 1](#_Toc188444083)

[2. Формулиране на проблема 1](#_Toc188444084)

[3. Описание на данните 1](#_Toc188444085)

[4. Методология 2](#_Toc188444086)

[4.1. Подготовка на данните 2](#_Toc188444087)

[4.2. Избор на модел 2](#_Toc188444088)

[4.3. Обучение на модела 3](#_Toc188444089)

[4.4. Оценка на модела 3](#_Toc188444090)

[4.5. Прогнози 3](#_Toc188444091)

[5. Бъдещи насоки 4](#_Toc188444092)

[5.1. Усъвършенствани модели 4](#_Toc188444093)

[5.2. Допълнителни характеристики 4](#_Toc188444094)

[5.3. Оптимизация на модела 4](#_Toc188444095)

[5.4. Прогнози в реално време 4](#_Toc188444096)

[6. Заключение 4](#_Toc188444097)

# **Въведение в проекта**

Целта на този проект е да се разработи усъвършенстван модел за прогнозиране на бъдещото електроенергийно потребление, базиран на задълбочен анализ на исторически данни. Тази прогноза има многоаспектно приложение, насочено към три ключови области: стратегическо планиране на енергийни ресурси, ефективно управление на съществуващата инфраструктура и оптимизиране на енергийното потребление. Проектът цели да предостави надежден инструмент за вземане на информирани решения в енергийния сектор, което ще допринесе за по-устойчиво и ефективно използване на електроенергийните ресурси.

# **Формулиране на проблема**

Електроенергийното потребление е повлияно от множество взаимосвързани фактори. Основните включват климатичните условия (температура и сезонност), демографски показатели (растеж на населението и урбанизация), макроикономически фактори (индустриално развитие и пазарни тенденции), технологични промени (нови технологии и енергийна ефективност), както и социално поведение (потребителски навици).

За анализ на тази комплексна система, проектът използва интегриран подход, съчетаващ анализ на исторически данни със съвременни статистически методи.

# **Описание на данните**

В рамките на този проект се използват симулирани данни, които са конструирани да отразяват максимално точно реалното електроенергийно потребление в град София. Времевият обхват на данните се простира върху десетгодишен период (от 2014 до 2024 година), като времевата резолюция е месечна. Те имат следните характеристики:

* *Дата*: Специфичният месец и година;
* *Електроенергийно потребление (MWh)*: Общото електропотребление в София през месеца;
* *Средна температура (°C)*: Средната температура за всеки месец;
* *Цена на електроенергията (BGN/MWh)*: Средната цена на електроенергията за месеца;
* *Празничен период:* Бинарна характеристика, която показва дали месецът е празничен (1 за август и декември, 0 в противен случай). Тази характеристика може да варира, според държава и религия.
* *Население (в хиляди):* Населението на София през месеца.

# **Методология**

## **Подготовка на данните**

### Зареждане на данните: Данните се зареждат от .csv файл

### Предварителна обработка:

### Конвертиране на колоната Date в обект от тип datetime;

### Обработка на липсващи стойности чрез метода на попълване „напред“ (forward fill);

### Извличане на допълнителни времеви характеристики като година и месец.

### Значение на характеристиките на данните:

### Добавените характеристики като средна температура, цени на електроенергията и празнични периоди присъстват, защото са основни фактори, влияещи на електропотреблението;

### Данните за населението има за цел да отразят дългосрочните тенденции.

## **Избор на модел**

### Линейната регресия беше избрана като модел за този проект, поради следните причини:

### *Простота*:

### Линейната регресия е лесна за имплементация и интерпретация;

### Тя предоставя информация за това как всяка характеристика влияе на електропотреблението.

### *Подходящ модел за непрекъсната променлива:*

### Електроенергийното потребление е непрекъсната променлива, която я прави подходяща за регресионни техники.

### *Връзка между характеристиките и целевата променлива:*

### Първоначалното изследване на данните показва сравнително линейна зависимост между целевата променлива и характеристиките като температура, население и цена на електроенергията.

### *Базов модел:*

### Линейната регресия служи като силен базов модел за бъдещи сравнения с по-усъвършенствани модели, ако е необходимо.

## **Обучение на модела**

### Разделяне на характеристики и целева променлива

### *Входни характеристики*: Year, Month, Avg\_Temperature\_C, Electricity\_Price\_BGN\_per\_MWh, Holiday\_Period, Population\_Thousands ;

### *Целева променлива*: Electricity\_Consumption\_MWh .

### Разделяне на данните

### *Данните са разделени на обучаващ (80%) и тестов (20%) набор*, за да се оцени представянето на модела върху невиждани данни.

### Обучение

### Моделът на линейна регресия е обучен чрез библиотеката Scikit-learn;

### Изчислени са коефициентите и пресечната точка, които представляват връзката между характеристиките и целевата променлива.

## **Оценка на модела**

### Метрики за оценка

### *Средна квадратична грешка (MSE)*: Измерва средната квадратична разлика между предсказаните и реалните стойности;

### *Корен от средна квадратична грешка (RMSE)*: Осигурява по-ясно интерпретация на грешките в същата единица като целева променлива;

### *Коефициент на детерминация (R²)*: Показва доколко моделът обяснява вариациите в данните.

### Резултати

### Представянето на модела върху тестовия набор предоставя базова линия за оценка на точността и надеждността му.

## **Прогнози**

### Обученият модел е приложен за генериране на прогнози за бъдещото електропотребление, като конкретен пример е януари 2025 г. За създаването на прогнозата моделът използва набор от ключови входни параметри: очакваната средна температура за периода, прогнозната цена на електроенергията, наличието на празнични дни и прогнозираната демографска динамика на населението.

# **Бъдещи насоки**

Настоящият модел, базиран на линейна регресия, предоставя солидна основа за прогнозиране на електропотреблението. Въпреки това, за повишаване на точността и обхвата на прогнозите, могат да бъдат разгледани следните потенциални подобрения:

## **Усъвършенствани модели**

### Изследване на нелинейни модели (напр. полиномна регресия или невронни мрежи) за улавяне на сложни зависимости;

### Използване на времеви модели като ARIMA или LSTM за улавяне на времеви зависимости.

## **Допълнителни характеристики**

### Включване на метеорологични данни (напр. валежи, снеговалежи, влажност);

### Добавяне на икономически индикатори като БВП или промишлено производство.

## **Оптимизация на модела**

### Извършване на мащабиране на характеристиките (напр. Ridge или Lasso регресия), за да се подобри обобщаването на модела.

## **Прогнози в реално време**

### Качване на модела като API за предоставяне на прогнози за електропотреблението в реално време.

# **Заключение**

Настоящият проект демонстрира успешна разработка и имплементация на прогностичен модел, базиран на линейна регресия, за предвиждане на електроенергийното потребление в град София. Моделът интегрира комплексен набор от влияещи фактори, включително климатични условия (средна температура), демографски показатели (население) и икономически параметри (цена на електроенергията), което позволява задълбочен и многоаспектен анализ на динамиката в енергийното потребление.

Постигнатите резултати, отразени в метриките за оценка на модела, потвърждават неговата приложимост за реални прогностични задачи в енергийния сектор. Същевременно, анализът на представянето разкрива специфични области, в които моделът може да бъде допълнително усъвършенстван, което открива път за бъдещи подобрения и оптимизации.