

Резюме на дипломна работа

“Система за прогнозиране на замърсяване на въздуха в София”

Замърсяването на въздуха е един от най-значимите екологични проблеми в съвременния свят, като носи сериозни рискове за човешкото здраве и околната среда. В дипломната работа се изследва процесът на прогнозиране на качеството на въздуха чрез прилагане на методи за машинно обучение и анализ на данни. Основната цел е разработването на интегрирана система, която събира, обработва, съхранява и анализира данни от сензорни устройства, за да предоставя точни прогнози за нивата на концентрация на замърсителите в атмосферния въздух.

Работата започва с анализ на съществуващите подходи и технологии за мониторинг на въздуха и прогнозиране на концентрациите на замърсителите, като се извършва сравнителен анализ на различни модели и алгоритми за прогнозиране, включително физически, статистически и модели на машинното обучение. След това се преминава към задълбочен сравнителен анализ на различните видове технологии за извличане, обработка и съхранение на данните, както и платформи за хостване на инфраструктура и разгръщане на модели за изкуствен интелект.

Следващата стъпка е детайлен анализ на изискванията към системата, който включва както функционални, така и качествени изисквания. Функционалните изисквания обхващат способността на системата да събира големи данни от динамична сензорна мрежа, да ги обработва и съхранява в сигурна среда, да прилага модели за машинно обучение и да предоставя резултатите чрез удобен интерфейс. Нефункционалните изисквания включват мащабируемост, сигурност, надеждност и ефективност на обработката, за да се гарантира стабилната работа на системата в реални условия.

След анализиране на изискванията се преминава към проектиране на системна архитектура, която включва модули за извличане, съхранение и обработка на големи обеми от данни, както и механизми за изграждане на модели за машинно обучение и инференция. Въз основа на анализа е избран LSTM (Long Short-Term Memory) модел. Освен основните слоеве, фокусирани върху обработката на потоци от данни, в системата са проектирани и спомагателни компоненти, които осигуряват поддръжка и оптимизация на основните работни процеси. Те включват механизми за оркестрация, мониторинг, сигурност, претрениране на модели и други функционалности, които допринасят за надеждността на цялостната система.

Реализацията на системата използва облачната платформа AWS, където данните се съхраняват и обработват чрез услуги като Redshift, S3 и Glue, а моделът за прогнозиране се изпълнява в SageMaker. Тестването на системата включва както модулно, така и системно тестване, за да се гарантира коректната работа на отделните компоненти и тяхното взаимодействие в цялостния работен поток.

В заключение, системата успява да изготвя прогнози относно нивата на основните замърсители, което я прави полезен инструмент за мониторинг и управление на качеството на въздуха. Възможностите за бъдещо развитие включват интегриране на допълнителни източници на данни, усъвършенстване на прогнозния модел и разширяване на системата с функционалности за прогнозиране на други екологични фактори като шумово замърсяване и качество на водата.