A blue logo with white text

Description automatically generated

**КУРСОВА РАБОТА**

**ПО ПРИЛОЖНО ИЗКУСТВЕН ИНТЕЛЕКТ**

**СПЕЦИАЛНОСТ: ИНФОРМАТИКА И СОФТУЕРНИ НАУКИ**

**НА ТЕМА: Business Intelligence - reporting tools and example(s)**

Изготвил**:**

* **Денис Светославов Патронов , фк:471222054**

**Business Intelligence - reporting tools and example(s)**

**Въведение**: Business Intelligence (BI) е процесът на събиране, анализ и представяне на данни, за да се подпомогне вземането на информирани бизнес решения. BI инструментите за отчитане (**reporting tools**) са софтуерни приложения, които извличат данни от различни източници, анализират ги и генерират лесно разбираеми отчети, графики и табла за управление.

**Идея:** Идеята е да се създаде модел, който предсказва печалбите за определен период от време. Той може да бъде изключително ценен за една компания, особено ако използва BI инструменти за анализ и визуализация на резултатите. Моделът би могъл да се интегрира с BI системи, за да предостави на компаниите прогнозни анализи, с лесно разбираем интерфейс, които подпомагат планирането и стратегическото управление.

**Алгоритми и инструменти:**

* **Random Forest Regressor**
* **Какво е?**  
  Random Forest е метод за ансамблово обучение, използван за регресия и класификация. Той работи чрез създаване на множество решаващи дървета по време на обучението и комбинира техните предсказания.
* **Как действа?**  
  При регресия, Random Forest предсказва стойността като взема средното (mean) на всички предсказания от отделните дървета.
* **Предимства:** 
  + Подобрява точността: Комбинацията от множество дървета намалява грешките.
  + Контролира пренасищането (overfitting): Индивидуалните дървета може да са склонни към overfitting, но Random Forest като цяло не е.
* **Pandas**
* **Какво е?**  
  Библиотека за Python, използвана за обработка и анализ на данни.
* **Основни характеристики:** 
  + DataFrames: Двуизмерна структура от данни, подобна на електронна таблица, която позволява удобно управление на големи обеми информация.
  + Инструменти за анализ: Лесни методи за сортиране, групиране, обработка на липсващи данни и много други.
* **Joblib**
* **Какво е?**Python библиотека за ефективна сериализация на обекти (запазване и зареждане на Python обекти като модели, данни и др.).
* **Използване в машинното обучение:** 
  + Запазване на обучени модели, за да се избегне повторното обучение.
  + Зареждане на модели за бързи прогнози.
* **Scikit-learn**
* **Какво е?**Библиотека за машинно обучение, която предлага инструменти за анализ на данни и моделиране.
* **Какво включва?** 
  + Алгоритми: Random Forest, SVM, KNN и др.
  + Инструменти за моделиране: Разделяне на данни, избор на характеристики, оценка на модели и др.
* **Защо е полезна?**Лесна за употреба и широко използвана от разработчици и изследователи.
* **Mean Squared Error (MSE)**
* **Какво е?**Средната квадратична грешка е мярка за разликата между действителните стойности и прогнозите на модела.
* **Формула**

A black background with white numbers and symbols

Description automatically generated

* + *y*i: Реални стойности.
  + *y*i^​: Прогнозирани стойности.
  + *n*: Брой на наблюденията.
* **Какво показва?**По-ниска стойност означава по-добра точност на модела.
* **R2 Score (коефициент на детерминация)**
* **Какво е?**Показва каква част от вариацията в целевите данни може да бъде обяснена от модела.
* **Формула:**

A black background with white text

Description automatically generated

* + *y*ˉ: Средната стойност на действителните стойности.
* **Какво означава резултатът?** 
  + 1: Перфектно предсказание.
  + 0: Моделът не обяснява никаква вариация.
  + <0: Моделът е по-лош от случайно предсказание.

**Примерна задача за Business Intelligence - reporting tools and example(s)**

1. **Задача**: Да се предскаже печалбата на дадена компания, на основа на нейните приходи и разходи.
2. **Данни:**

* `Month`: Месец на записа.
* `Year`: Година на записа.
* `Income`: Приходиоте на компанията за даден месец.
* `Expenses`:Разходите на компанията за даден месец.
* `Number\_of\_Employees`:Брой служители в компанията през даден месец.
* `Profit`: Печалбата за компанията през даден месец.

1. **Обработка на данни**

**Зареждане на данни**

* Данните се зареждат от **CSV файл** с помощта на библиотеката **Pandas**.

**Предварителна обработка на данни**

* Данните се обработват предварително, за да се създадат следните нови колони:
  + **Month\_Year** (Месец\_Година)
  + **Actual\_Profit** (Реална печалба)
  + **Month** (Месец)
  + **Year** (Година)

**Дефиниране на характеристики и цел**

* **Характеристики (X):**
  + Определени са като **Приходи (Income)** и **Разходи (Expenses)**.
* **Цел (y):**
  + Определена е като **Печалба (Profit)**.

**Разделяне на данни**

* Данните се разделят на **обучаващ набор** и **тестващ набор** с помощта на функцията **train\_test\_split** от библиотеката **Scikit-learn**.

1. **Обучение и оценка на модела:**

**Обучение на модела**

* **Random Forest Regressor се обучава с помощта на GridSearchCV, за да се намерят най-добрите хиперпараметри.**

**Оценка на модела**

* **Моделът се оценява с помощта на:**
  + **Средна квадратична грешка (Mean Squared Error, MSE).**
  + **Коефициент на детерминация (R² Score).**

1. **Запазване на модела и прогнози:**

**Запазване на модела**

* **Обученият модел се запазва с помощта на библиотеката Joblib.**

**Създаване на прогнози**

* **Правят се прогнози върху тестовите данни.**

**Запазване на прогнозите**

* **Прогнозите се записват в CSV файл.**

**Резултати от експеримента за предксазване на печалбите в дадена компания:**

* Моделът постигна **Средна квадратична грешка (MSE)** от **X= 7908.277777777769** и **Коефициент на детерминация (R² Score)** от **Y = 0.9640532828282828** върху тестовите данни.
* **Заключение:** Моделът **Random Forest Regressor** успя да предскаже печалбата на компанията въз основа на нейните приходи и разходи с добра точност. Използването на **GridSearchCV** помогна за намирането на най-добрите хиперпараметри, което подобри производителността на модела. Резултатите показват, че моделът може да бъде използван за бъдещи прогнози на печалбите, като подпомага финансовото планиране и вземането на решения за компаниите, които го използват.
* **Резултат в конзолата:**



* **Резултат от предсказването, сравнен с реалната печалба, запазен във csv файл:**

A screenshot of a computer

Description automatically generated