



Курсова работа

Приложен изкуствен интелект

„Директна невронна мрежа за определяне риска при отдаване на кредит“

- изготвил: Аркан Ахмедов
- преподавател : Александър Ефремов

1. Въведение

Идея

Целта на проекта е да се разработи модел за машинно обучение за оценка на кредитния риск с помощта на невронни мрежи. Целта е да се предскаже вероятността от кредитно неизпълнение въз основа на различни финансови показатели. Тази задача е от решаващо значение във финансовия сектор за вземане на информирани решения за отпускане на кредити.

Терминология

Оценка на кредитния риск: Процесът на определяне на вероятността даден кредитополучател да не изпълни дълговите си задължения.

Невронна мрежа: Невронна мрежа: изчислителен модел, вдъхновен от начина, по който биологичните невронни мрежи в човешкия мозък обработват информация.

TensorFlow и Keras: Библиотеки с отворен код за машинно обучение и дълбоко обучение.

Подходи и методи

Подходът включва използването на захранваща невронна мрежа, реализирана с помощта на TensorFlow и Keras, обучена върху исторически финансови данни. Основните стъпки включват предварителна обработка на данните, обучение на модела и оценка.

2. Теория

Описание на алгоритмите

Напредваща невронна мрежа: Тип изкуствена невронна мрежа, при която връзките между възлите не образуват цикъл. Тази мрежа включва входен слой, скрити слоеве и изходен слой.

StandardScaler: Това е техника за предварителна обработка, която стандартизира характеристиките чрез премахване на средната стойност и мащабиране до единица дисперсия.

Изпълнение на основни инструменти за работа с данни

Pandas: Библиотека на Python, която се използва за манипулиране и анализ на данни. В този проект тя се използва за четене и предварителна обработка на набора от данни.

Разделяне на тренировъчни и тестови данни: Метод за оценка на ефективността на даден модел чрез разделяне на набора от данни на набор за обучение и набор за тестване.

3. Пример:

Описание на задачата

Задачата е да се изгради модел на невронна мрежа за оценка на кредитния риск въз основа на различни атрибути като баланс по сметка, размер на кредита, продължителност на трудовия договор и други.

Описание на данните

Наборът от данни съдържа финансови атрибути на физически лица, включително салда по сметки, суми на кредити и история на плащанията. Целевата променлива е "Кредитоспособност", която показва дали дадено физическо лице е просрочило кредит.

Представяне на малка част от подготвените данни. Моделът е трениран на 800 реда и тестван върху 200.

Creditability	Account_ID	Duration_Payment	Purpose	Credit_Amount	Value_Saved	Length_of_Instalment	Sex	Marital_Status	Guarantor	Duration_Months	Most_valued	Age_years	Concurrent	Type_of_a	No_of_Credit	Occupation	No_of_dependents	Telephone	Foreign_Worker
1	1	18	4	2	1049	1	2	4	2	1	4	2	21	3	1	1	3	1	1
1	1	9	4	0	2799	1	3	2	3	1	2	1	36	3	1	2	3	2	1
1	2	12	2	9	841	2	4	2	2	1	4	1	23	3	1	1	2	1	1
1	1	12	4	0	2122	1	3	3	3	1	2	1	39	3	1	2	2	2	1
1	1	12	4	0	2171	1	3	4	3	1	4	2	38	1	2	2	2	1	2
1	1	10	4	0	2241	1	2	1	3	1	3	1	48	3	1	2	2	2	1
1	1	8	4	0	3398	1	4	1	3	1	4	1	39	3	2	2	2	1	2
1	1	6	4	0	1361	1	2	2	3	1	4	1	40	3	2	1	2	2	1
1	4	18	4	3	1098	1	1	4	2	1	4	3	65	3	2	2	1	1	1
1	2	24	2	3	3758	3	1	1	2	1	4	4	23	3	1	1	1	1	1
1	1	11	4	0	3905	1	3	2	3	1	2	1	36	3	1	2	3	2	1
1	1	30	4	1	6187	2	4	1	4	1	4	3	24	3	1	2	3	1	1
1	1	6	4	3	1957	1	4	1	2	1	4	3	31	3	2	1	3	1	1
1	2	48	3	10	7582	2	1	2	3	1	4	4	31	3	2	1	4	1	2
1	1	18	2	3	1936	5	4	2	4	1	4	3	23	3	1	2	2	1	1
1	1	6	2	3	2647	3	3	2	3	1	3	1	44	3	1	1	3	2	1

Обработка на данните

Етапите на обработка на данните включват:

Зареждане на данните с помощта на Pandas.

Разделяне на характеристиките и целевата променлива.

Мащабиране на характеристиките с помощта на StandardScaler.

Моделиране на

Създава се трислойна невронна мрежа с помощта на Keras. Мрежата се състои от два скрити слоя с ReLU активация и изходен слой със сигмоидна активационна функция за двоична класификация.

4. Експеримент

Резултати

Моделът е обучен върху набор от данни и е оценен за неговата точност. Той прогнозира дали новите лица ще просрочат заемите си въз основа на техните финансови профили.

Анализ

Точността на модела се изчислява върху тестовите данни.

Направени са прогнози върху нови, невиджани данни, като се използва обученият модел.

Изводи

Експериментът демонстрира възможностите на невронните мрежи за оценка на кредитния риск. Ефективността на модела се измерва количествено с помощта на точността и той дава обещаващи резултати за практически приложения при вземането на финансови решения.

Точността с която моделът предсказва тестови примери.

Test accuracy: 0.7599999904632568

Заклучение

Този проект илюстрира ефективността на невронните мрежи при оценката на кредитния риск. Той подчертава значението на предварителната обработка на данните и силата на TensorFlow и Keras при изграждането на прогнозни модели. По-нататъшните подобрения биха могли да включват настройка на хиперпараметри и използване на по-сложни мрежови архитектури.