Пловдивски университет „Паисий Хилендарски”

**Факултет по математика и информатика**

**Катедра *Компютърна информатика***

**ДИПЛОМНА РАБОТА**

**Прогнозиране на успеха на студенти чрез използване на методи за машинно обучение**

**Дипломант: Михаил Йорданов Йорданов**

**Фак. №: 1801322019**

**Научен ръководител: доц. д-р Силвия Гафтанджиева**

**Пловдив**

**2022 г.**

**Съдържание**

**TODO:**

**(hyperlinked content viewer)**

1. УВОД
2. МАШИННО ОБУЧЕНИЕ
   * 1. ДЕФИНИЦИЯ
     2. ОБЩ МОДЕЛ НА МАШИННОТО ОБУЧЕНИЕ
     3. ПРОГНОЗНИ АНАЛИЗИ
3. МЕТОДИ
   * 1. ИЗБРАНИ МЕТОДИ
        1. GRADIENT BOOSTED
        2. RANDOM FOREST
        3. EXTRA TREES
        4. POLYNOMIAL REGRESSION
        5. LINEAR REGRESSION
        6. ELASTICNET REGRESSION
        7. BASELINE
4. МАТЕРИАЛИ
5. ПРИЛОЖЕНИЕ И РЕЗУЛТАТИ
6. ОЦЕНКА
   * 1. СРАВНЕНИЕ НА МЕТОДИ
7. ЗАКЛЮЧЕНИЕ
8. СПИСЪК НА СЪКРАЩЕНИЯТА
9. ЛИТЕРАТУРА

**1. Увод**

Бързото развитие и широкото разпространение на компютърните системи и Интернет водят до разрастването на публично достъпни данни, които да бъдат анализирани. Тези данни могат да бъдат под формата на всякакъв вид информация за определена дейност или статистика, като например трафик на уеб сайт, потребителски предпочитания и навици или поръчки. Основният проблем често е размерът на данните, тъй като те са твърде големи, за да бъдат анализирани от хора. Компютрите от друга страна могат да обработват големи количества от данни много бързо.

Концепцията за машинно обучение се поражда от тази обстановка, като нейната идея е “изкуствено” генериране на знания от опит: системата се учи от примери и може да ги обобщи след завършване на етапа на обучение.

Машинното обучение намира приложение в доста сфери от нашето ежедневие. Например при онлайн търсачките като Google, Bing, DuckDuckGo и много други, се използва машинно обучение за осъществяване на по-добра връзка между търсените фрази и показаните като резултат уеб страници. Анализирайки съдържанието на уеб страниците, търсачките могат да дефинират кои думи и фрази са най-значителни при дефинирането на конкретна уеб страница. Използвайки това връщат най-уместните резултати за дадена търсена фраза.

В зависимост от вида на данните, алгоритмите за машинно обучение се разделят на два вида – контролирано (supervised) и безконтролно (unsupervised) машинно обучение. При контролираното обучение ние “учим” модела и след това с тези знания той може да прогнозира непознати или бъдещи случаи. Много важен отличителен белег на контролираното обучение е, че данните са белязани (labeled). От своя страна контролираното обучение се дели на 2 вида: Класификация и Регресия. Като Класификацията е процес, при който се прогнозират дискретни категории или означения, а при Регресията се прогнозират непрекъснати стойности.

***Тематиката на дипломната работа*** е свързана със сравняването на различни методи за машинно обучение, които прогнозират крайния успех на студенти. Оценява се точността на методите, както и потенциалната грешка в прогнозата.

**2. Машинно обучение**

2.1. Дефиниция

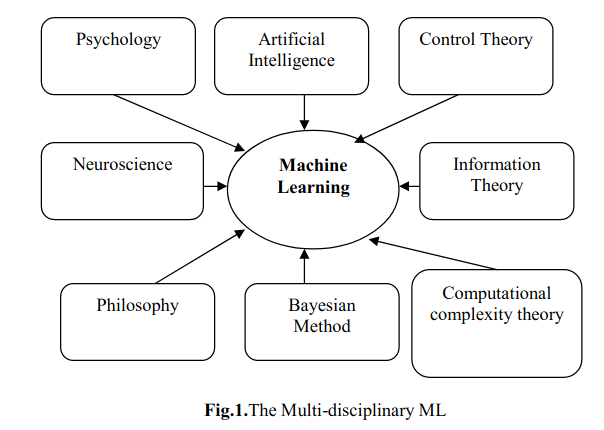
Ученето е процес на придобиване или модифициране на вече съществуващи стойности, познания, предпочитания или поведения. Машините, за разлика от хората, зависят от данни: учат се от опит. На фундаментално ниво машинното обучение е категория от изкуствения интелект, която позволява на компютрите да “мислят” и да се самообучават. По този начин те променят действията си, с цел подобряване на точността, която се измерва от гледна точка на това колко пъти взетото решение е довело до правилния резултат.

“Казва се, че компютърната програма се учи от опит *E* със зависимост от клас от задачи *T* и производителност *P*, ако производителността за извършването на задачите *T*, както е измерена от *P*, се увеличава с опита *E*.”

* Том М. Мичъл

С други думи, машинното обучение е възможността на компютъра да се учи от опит, който обикновено е даден във формата на входни данни. Гледайки данните, той може да намери зависимости в тях, които са твърде сложни, за да бъдат намерени от човек. Машинното обучение може да се използва за откриване на скрити структури от класове в неструктурирани данни или за откриване на зависимости в структурирани данни, за да направи прогноза.

Машинното обучение е мултидисциплинарно поле, имайки широк обхват от изследователски области, подпомагащи неговото развитие (фиг. 1). ###### edit photo

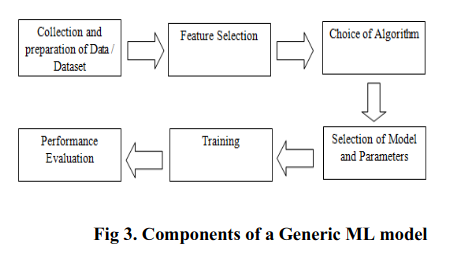


2.2 Общ модел на машинното обучение

Машинното обучение се използва за решаването на проблеми, чрез самообучение от страна на машината. Такива проблеми имат 3 особености:

* Клас от задачи
* Подобряване на производителността
* Процесът на трупане на опит

Общият модел на машинното обучение се състои от 6 компонента, независещи от възприетия алгоритъм (фиг. 2). ### TODO EDIT PHOTO



Всеки компонент има за цел да изпълни конкретна задача:

1. Събиране и подготовка на данни: Основната задача в този процес е да се съберат и подготвят данните във формат, който може да се подаде като вход на алгоритъма.
2. Избор на функции: —————
3. Избор на алгоритъм: ————-
4. Избор на модел и параметри: ————-
5. Трениране: ————
6. Оценка на производителността: —————

**3. Методи**

Текст

**4. Материали**

Текст

**5. Приложение и резултати**

Текст

**6. Оценка**

Текст

**7. Заключение**

Текст

**8. Списък на съкращенията**

Текст

**9. Литература**

Текст