

**MINISTRY OF HIGHER EDUCATION
KABUL UNIVERSITY
IS DEPARTMENT**

**Mobile price range prediction
Naïve Bayes Algorithm**

Instructor: Prof. Ziaullah Momand

Subject: Data science

Group members: Abdul Bari Amiri & Ibrahim Zazai

Business understanding

دیتا ساینس یکی از جمله علم های معاصر میباشد که در تمام ابعاد زندگی و جهت پیشبرد امورات زندگی از آن استفاده می شود. قسمی که ما آموختیم که دیتا ساینس علم است که بخاطر تحلیل سروی نفوذ شماری و آن عده مشکلات که در دنیای حقیقی میباشد بحث نموده و یک راه حل پیدا میکند.

مهمترین عنصر در این علم دیتا میباشد که دیتا عنصر اساسی مودل های میباشد که جهت پیش بینی و انالیز کردن دیتا استفاده می شود پس شناختن دیتا در مودل جز عمده و اساسی میباشد پس در این پروژه که مد نظر گرفتیم ما نخست در مورد دیتای که برای مودل ما ضرورت است با مراجعه به چند شرکت تولید و فروش موبایل ارقام و مشخصات که بخاطر ارایه و آماده کردن مودل ضرورت است جمع آوری نمودیم

Data understanding

قسمی که قبلا گفتیم دیتا جز عمده و اساسی این علم بوده و باید هر data scientist در مورد دیتا معلومات داشته باشد. دیتا های جمع آوری شده از طریق تحقیق و سروی که داشتیم بدست آمده و تمام آن مشخصات که در مورد پروژه ضروری بوده جمع آوری کردیم .

Dataset پروژه ذیل از وب سایت (www.kaggle.com) گرفته شده و دیتای مودل که ما ساختیم در مورد (Mobile price prediction) بوده و این دیتا مودل دارای به تعداد ستون میباشد.

(Missing value) که یک جز اساسی و مهم در زمان جمع آوری دیتا میباشد اگر ما هر قدر در مودل خود کمترین (Missing value) داشته باشیم میزان (performance) مودل ما بالا میرود و (accuracy) بهتر میداشته باشد. در دیتا مودل ذیل تلاش بر این شده تا از (Missing value) ها جلوگیری شود و در دیتا مودل ذیل ما هیچ نوع (Missing value) نداریم.

Data analysis

در data science ما چندین الگوریتم را شناختیم و آموختیم که هر کدام آن موثریت و بعضی ها نواقص را دارد پس در دیتا مودل که ساختیم ما از الگوریتم Naïve bays استفاده نمودیم که مودل ما را به سه کتگوری 1 که به معنی قیمت پایین 2 به معنی قیمت متوسط و 3 یعنی قیمت بالا predict می کند.

Mobile range prediction using (Naïve bays Algorithm)

Mobile Price Range Prediction.

Abstract

There is a new mobile company and the CEO wants to give a tough fight to big companies like Apple, Samsung etc. He has no idea of how to estimate the price of

mobiles his company creates. A simple assumption of the prices will not be profitable in this competitive world. To solve this problem, he collects sales data of mobile phones of various companies. Thus, the CEO has hired you to find out some relation between features of a mobile phone (eg:- RAM, Internal Memory etc) and its selling price

Problem Statement:****

Predict a price range, indicating how high the price is, using the Naive Bayes algorithm.

Dataset Information:***

- battery power: Total energy a battery can store in one time measured in mAh
- clock speed: The speed at which microprocessor executes instructions
- fc: Front Camera megapixels
- int_memory: Internal Memory in Gigabytes
- m_dep: Mobile Depth in cm
- mobile_wt: Weight of the mobile phone
- n_cores: Number of cores of a processor
- pc: Primary Camera megapixels
- px_height: Pixel Resolution Height
- px_width: Pixel Resolution width
- ram: Random Access Memory in MegaBytes
- sc_h: Screen Height of mobile in cm
- sc_w: Screen Width of mobile in cm
- talk_time: The longest time that a single battery charge will last when you are
- price_range: This is the target variable with the value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).

Scope:****

- Prepare and analyses data.
- Train Naive Bayes model with data and check it's performance.

Learning Outcome:****

The company will get a better understanding of how the variables are linked to each other and how the EDA approach will help them gain more insights and knowledge about the data that we have and classify the data into similar groups using Naive Bayes. They will also learn about how to choose important features using Random Forest.