



دانشگاه شهید بهشتی

دانشکده علوم پایه

# پروژه درس یادگیری ماشین

## Bank-Marketing

استاد: دکتر حسین حاجی ابوالحسن

دانشجو: محمد عضدی

مقطع: کارشناسی ارشد

گرایش: علوم داده ها

تیر ماه ۱۳۹۹

# **فهرست :**

**مقدمه و طرح مسئله**

**تاریخچه**

**شرح داده ها**

**شرح روش**

**بررسی نتایج**

**منابع**

## مقدمه و طرح مسئله :

کمپین های بازار بانکی امروزه بیشتر به عقاید انسان متخصص در مورد انتخاب مشتریان بالقوه متکی هستند. این روش وقت گیر است و فاقد دقت است. همانطور که بانک ها اطلاعات بسیار و معامله های مشتری ساختار یافته و دقیق سوابق در دست دارند ، مطلوب است برای تصمیم گیری مبتنی بر داده ها سیستم هایی با ضریب موفقیت بالا در تبلیغات ساخته شود.

مدل ها و تکنیک های یادگیری ماشین پتانسیل بسیار خوبی برای نشان دادن قدرت خود در چنین تنظیمات مشکل را دارند. این گزارش پروژه نشان می دهد که الگوریتم های یادگیری ماشین چگونه می توانند بخصوص در چنین تنظیمات مشکل بصورت عملی استفاده شود.

این پروژه در مورد اجرای روش های یادگیری ماشین بر روی داده هاست(داده ها از سایت [kaggle.com](https://www.kaggle.com) گرفته شده است). در این پروژه مجموعه داده (Data set) در نظر گرفته شده شامل داده های یک بانک در کشور پرتغال است که یک کمپین جدید جهت سرمایه گذاری راه اندازی کرده و قصد دارد با توجه به داده هایی که از افراد دارد پیشگویی کند:

**که آیا افراد در این کمپین شرکت می کنند یا خیر.**

اطلاعات مربوط به فعالیتهای بازاریابی مستقیم یک موسسه بانکی پرتغال است. کارزارهای بازاریابی مبتنی بر تماسهای تلفنی بود. اغلب ، بیش از یک تماس با همان مشتری لازم بود ، برای دسترسی به این محصول ("بله") یا نه ("نه") مشترک اگر دسترسی داشته باشید.

## تاریخچه:

چشم انداز تاریخی از بازاریابی و تکامل فناوری اطلاعات را در این صنعت فشرده اطلاعاتی ارائه می دهد ، راه هایی را برای تحقیقات آینده که از این منظر ناشی می شود ، پیشنهاد می کند. با استفاده از یادگیری ماشین در ده سال اخیر، بانکها توانایی پیشگویی برای جذب مشتریان بالقوه در بانکها را بدست آورده است. از جمله استفاده از شبکه های عصبی و مدل های کارآمد یادگیری ماشین در بازاریابی امور بانکی باعث کم شدن هزینه های تبلیغات و سرعت بیشتر شده است. اکثر بانکهای موفق امروزی از این روشهای داده کاوی و پردازش داده ها استفاده می کنند که با درصد درستی بالایی همراه است.

## شرح داده ها:

این داده شامل ۲۱ ویژگی (ستون) و ۴۱۱۸۸ مشاهده (سطر) است. که مجموع کل داده ۴۱۱۸۸ است و حال ویژگی ها را اندکی توضیح می دهیم:

مهمترین داده های ما اعم از سن، شغل، وضعیت تاهل (طلاق-ازدواج-مجرد-مشخص نیست)، آموزش (بیسواد- ابتدایی-متوسطه-حرفه-دانشگاه-مشخص نیست)، وام، وام مسکن، وام شخصی است.

یک سری از داده ها که مرتبط با آخرین تماس با کمپین فعلی:

تماس: نوع ارتباط با مخاطب (طبقه بندی شده: "تلفن همراه"، "تلفن")

ماه: آخرین ماه تماس سال (طبقه بندی شده: 'jan', 'feb', 'mar', ..., 'nov', 'dec')

روز: آخرین روز تماس هفته (طبقه بندی شده: 'mon', 'سه', 'wed', 'thu', 'fri')

مدت زمان: آخرین مدت زمان تماس ، چند ثانیه (عددی).

نکته مهم: این ویژگی به شدت بر روی هدف خروجی تأثیر می گذارد (به عنوان مثال ، اگر مدت زمان = ۰ باشد ، 'y = 'no'). با این وجود ، مدت زمان قبل از انجام تماس مشخص نیست. همچنین

، پس از پایان تماس ۷ کاملاً مشخص است. بنابراین ، این ورودی فقط باید برای اهداف معیار گنجانده شود و در صورتی که قصد داشتن یک مدل پیش بینی واقع گرایانه باشد ، باید از آن صرف نظر کرد.

## شرح روش :

در طی این پروژه که در فضای ژوپیتتر نوت بوک انجام میشود، ابتدا یک سری دیداری کردن و پیش پردازش داده ها را انجام میدهم که بیشتر به صورت نمودارهای فراوانی نشان داده میشود. و در ادامه راستای کاهش بعد قدم برمیداریم.

با بررسی بیشتر داده ها و متغیرها، تصمیم به حذف یا اضافه کردن بعضی از متغیرها میگیریم و در نهایت با اعمال مدل‌های زیر بروی داده ها میپردازیم:

### 1: Logistic regression

### 2: Decision tree

### 3: Random Forest

## بررسی نتایج:

هدف از انتخاب مدل ، یافتن معماری شبکه با بهترین خصوصیات تعمیم ، یعنی مواردی است که خطاهای موجود در موارد انتخاب شده مجموعه داده را به حداقل می رساند. در پایان پروژه بعد از اعمال مدل‌های بالا و خروجی گرفتن از داده برای انتخاب بهترین مدل خود برای بررسی و پیش بینی داده هایمان استفاده میکنیم. که در نهایت مدل random forest بهترین مدل انتخابی ما خواهد بود.

## منابع:

Understanding Machine Learning: From Theory to Algorithms, by Shai Shalev-Shwartz and Shai Ben-David

Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd Edition) by Aurelien Geron

Bank Marketing with Machine Learning-Zewei Chu-April 19, 2015

link:

<https://www.kaggle.com/henriqueyamahata/bank-marketing>

Dataset from : <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>

<https://www.neuraldesigner.com/learning/examples/bank-marketing-campaign>

[https://www.researchgate.net/publication/340788220\\_Mining\\_a\\_Marketing\\_Campaigns\\_Data\\_of\\_Bank](https://www.researchgate.net/publication/340788220_Mining_a_Marketing_Campaigns_Data_of_Bank)

<https://nycdatasience.com/blog/student-works/machine-learning/machine-learning-retail-bank-marketing-data/>