**به نام خدا**

# گزارش پروژه مقایسه سه الگوریتم لاجستیک رگرسیون و نزدیکترین همسایه مشترک و ناوی بیز

**درس داده کاوی**

**دانشگاه آزاد واحد تهران جنوب**

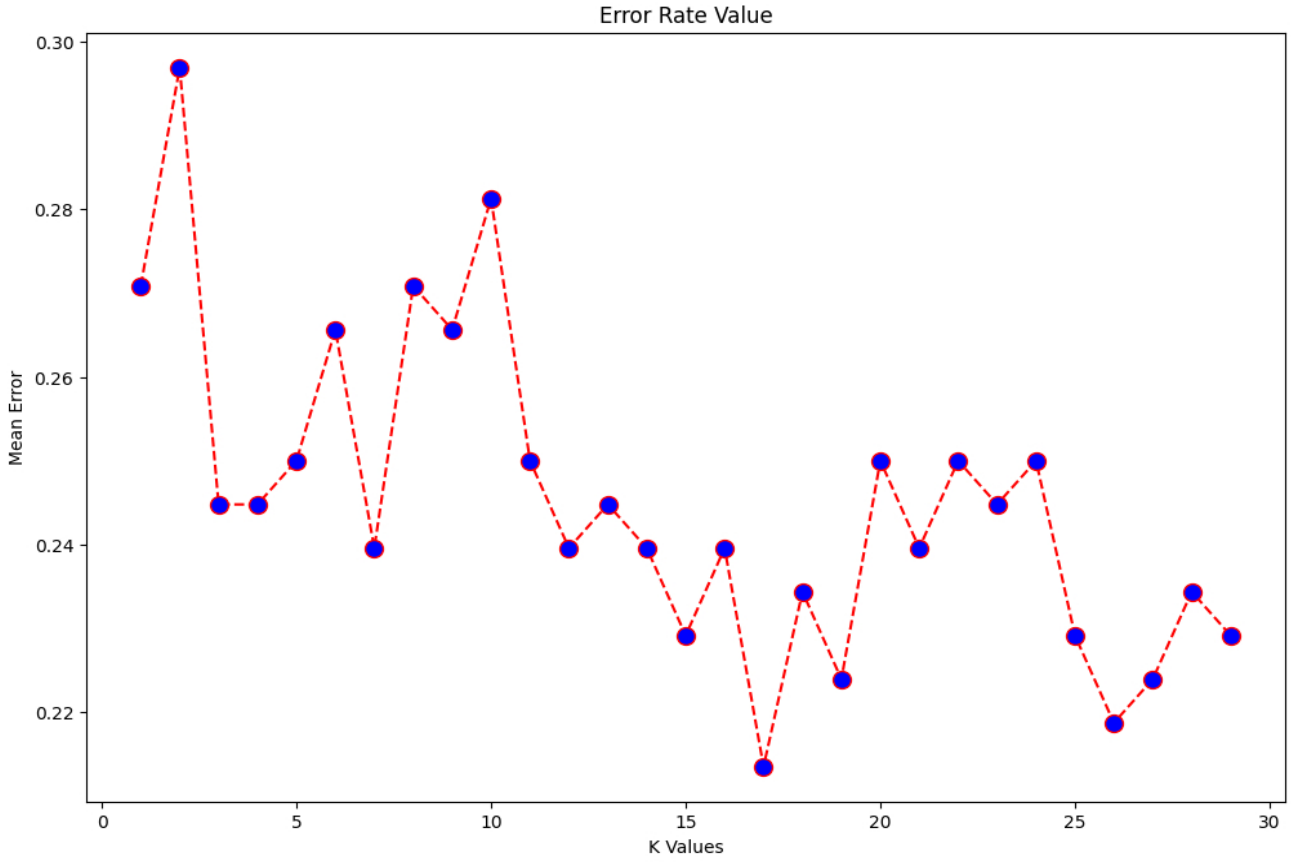
**سید محسن شکرابی**

# پیاده‌سازی سه مدل پیشبینی

در این پروژه دیتاستی بر اساس مشخص کردن دیابت افراد براساس یکسری ویژگی تشکیل شده‌است.

درابتدا پس از لودکردن فایل csv ۲۵ درصد داده ها را به عنوان داده Test از داده Trainجدا میکنیم و سپس به کمک تابع StandardScaler() پیش‌پردازشی روی داده‌ها جهت مقیاس‌پذیری داده‌ها صورت میپذیرد.

در مرحله بعدی با فراخوان کتابخانه LogisticRegression پیشبینی روی این مدل انجام میدهیم و همچنین در مرحله بعدی با استفاده از کتابخانه KNeighborsClassifier یادگیری را روی این مدل به ازای k بین ۱ تا ۳۰ انجام میدهیم و همانطور که از شکل زیر مشخص است به ازای k = 17 کمترین خطا را دارد.



**سپس به کمک کتبخانه GaussianNB پیشبینی روی این مدل انجام میدهیم.**

# محاسبه دقت هر مدل

**در این مرحله به کمک تابع metrics.accuracy_score دقت هر مدل را براساس درصد محاسبه میکینیم که بصورت زیر مشاهده میشود.**

**Logistic Regression: 0.78125,'**

**K Nearest Neighbor: 0.7864583333333334,**

**Naive Bayes: 0.765625**

## محاسبه کانفیوژن ماریس برای هر مدل

**در این قسمت با استفاده از تابع metrics.confusion_matrix این تابع را برای سه مدل انجام میدهیم که خروجی به شکل زیر قابل مشاهده است.**

**{'Logistic Regression': array([[111, 14],**

**[ 28, 39]]),**

**'K Nearest Neighbor': array([[112, 13],**

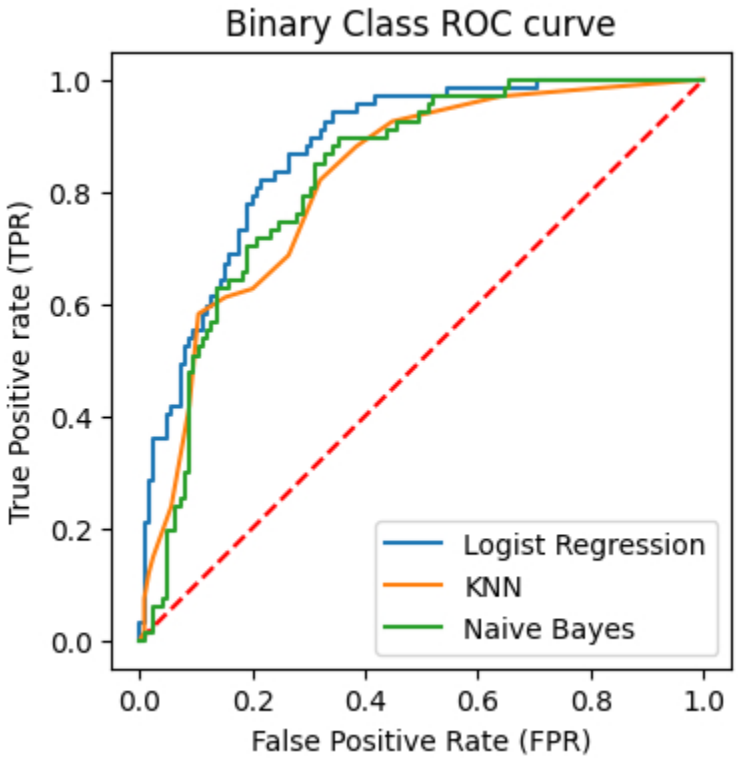
**[ 28, 39]]),**

**'Naive Bayes': array([[105, 20],**

{([[۴۲ ,۲۵ ]

## رسم نمودار ROC برای هر مدل

**در آخرین مرحله نمودار ROC برای هر مدل به شکل زیر رسم میشود.**




## نحوه اجرا کد در google colab

جهت اجرای کد در محیط گوگل کولب وارد لینک http://colab.research.google.com/github شوید و در قسمت سرچ گیت هاب آدرس گیت به نشانی https://github.com/mohsenshekarabi/homework2_datamining/ را سرچ کنید و فایل CompareModelsPrediction.ipynb را باز کنید.