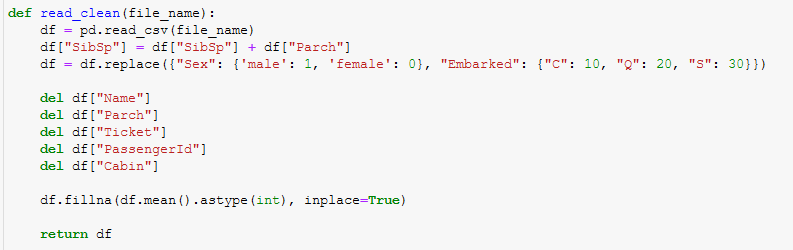
**شش.**

بعد از خواندن داده‌ها، شروع به پیش‌پردازش می‌کنیم:



داده‌های مهم غیرعددی را به داده‌های عددی map می‌کنیم، داده‌های بی‌ربط را حذف می‌کنیم و مقدارهای ناموجود را با میانگین بقیه داده‌ها جای‌گزین می‌کنیم.

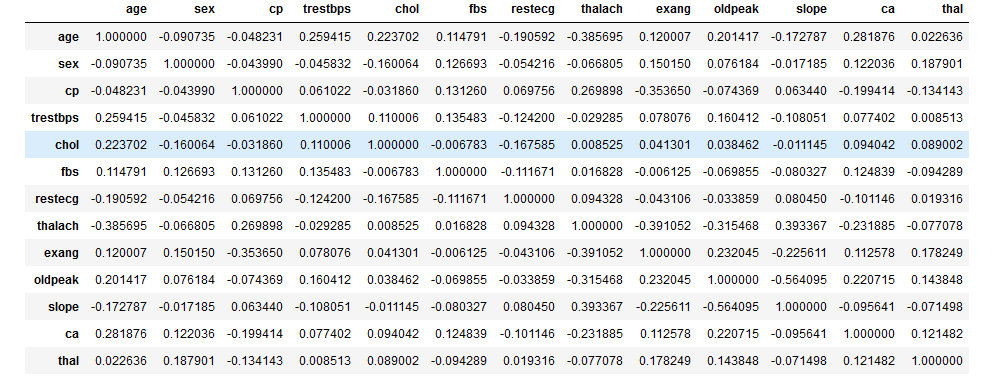
با معیارهای جینی و آنتروپی در عمق‌های ۲ – ۴ و ۵ تست کردم و بهترین جواب به دست آمده که با ملاک جینی و عمق ۵ بود، امتیاز ۷۷.۵ را از سایت Kaggle گرفت:



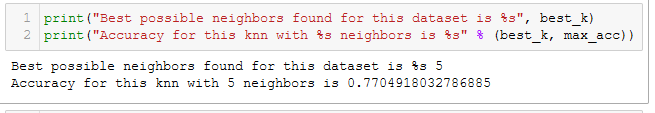
(با استفاده از RandomForest به جواب ۷۷.۷ رسیدم)

**هفت.**

با محاسبه‌ی Correlation ها به جدول زیر می‌رسیم که نشان می‌دهد ماکسیمم آن بین ستون‌های مختلف ۴۷ درصد است پس خیلی نمی‌توان کاهش بعد داد:



با ایتریت‌کردن روی k های مختلف برای knn به این می‌رسیم:



و با انجام نایو بیز چندجمله‌ای به نتیجه‌ی زیر:

