Do you think these predictions are good?

Many modern companies of different sizes and individuals should use predictions, otherwise people won't know what to expect and how to react to some events: weather changing, population changing, market satiation changing, etc. People are relying on modern technologies, so they be should provide effective data analysis and predictions. Data analysis technologies should be available for all kind of users whether they have programming experience or not. Models should be accurate, deployment should be quick and tools should be available for collaborative work. I like DataRobot solution because it provides effective data analysis for all kind of data, it's easy to use and it's scalable and reliable at the same time.

I can provide my opinion about Google Predictions API from a first time user point of view. I liked that a user can try it on-line without using any programming languages and using trial version of the API. In addition, it's really convenient that it has API for Python and other popular languages. Documentation a little bit tricky sometimes.

I decided to create a program to analyze if some message is a spam. Nowadays people get too much spam through text messages and emails. It's important to have a good spam detection filter.

I got a dataset with list of spam and ham messages. Then I created a project in Google Developers Console and uploaded the dataset. Using Python I created and trained a model in Google Prediction API. Then I checked if phrase: "Hey buy some stuff" was a spam. The result was that the phrase is more like ham message then spam.

Each steps results are provided in a screen shot in result-ps.png file. I also have a step for deletion a model, but it's commented for now in the code.

Chosen dataset is quite simple: it contains only two categories and two columns. For that data the model was trained good and provided acceptable prediction. I think for similar types of datasets it's efficient to use Google Prediction API and Python. The model was trained fast enough and provided a good result. I'm not sure if the API can be good for larger datasets. It can be slow during training stage and can consume too much bandwidth resources. In addition I'm not sure if the solution can be scalable and the model can be easy to deploy.