Valerii Podymov

Assignment, part 3

We have created an extension of the data set that describes 55 agents and the 34103 calls that they have placed. Some debtors were called by multiple agents before they repaid or their account was abandoned. Figure out what to suggest to the collection agency, like you would if you work with us. This is purposely an open-ended question so we can see how you would approach the challenge: write a short essay about how you would tackle this challenge and implement a proof-of-concept script and/or visualization using the language and/or tools of your choice.

Suggested Approach to the Problem

According to the conditions mentioned above, I would like to stick with the following plan:

1. Understanding the available data sets.

At first glance we have additional dimensions to take into account:

- Basic characteristics of placed calls (time of day, day of week, duration etc.)
- Actual content of the calls., i.e. speech and voice of agents and clients

Here I would like to emphasize the role of both speech and voice. The role of Speech Analytics is to provide the content of a certain conversation by transcribing recorded calls.

Voice Analytics focuses on the emotional aspect of speech, or how people speak in a conversation. I would not underestimate the importance of voice analysis (see for example studies by Prof. A. Mehrabian on paraverbal messages), since it allows assessing the emotional state of the speaker and avoiding confusion over the meaning of a given word or phrase used.

2. Feature selection

I addition to the initial data set, I would use basic characteristics of the placed calls and features extracted from transcribed speech (using NLP techniques) and voice (syllable emphasis, tone, pitch etc.) to figure out how to define if a debtor tends to behave collaboratively. And finally, to identify factors that make the phone calls successful.

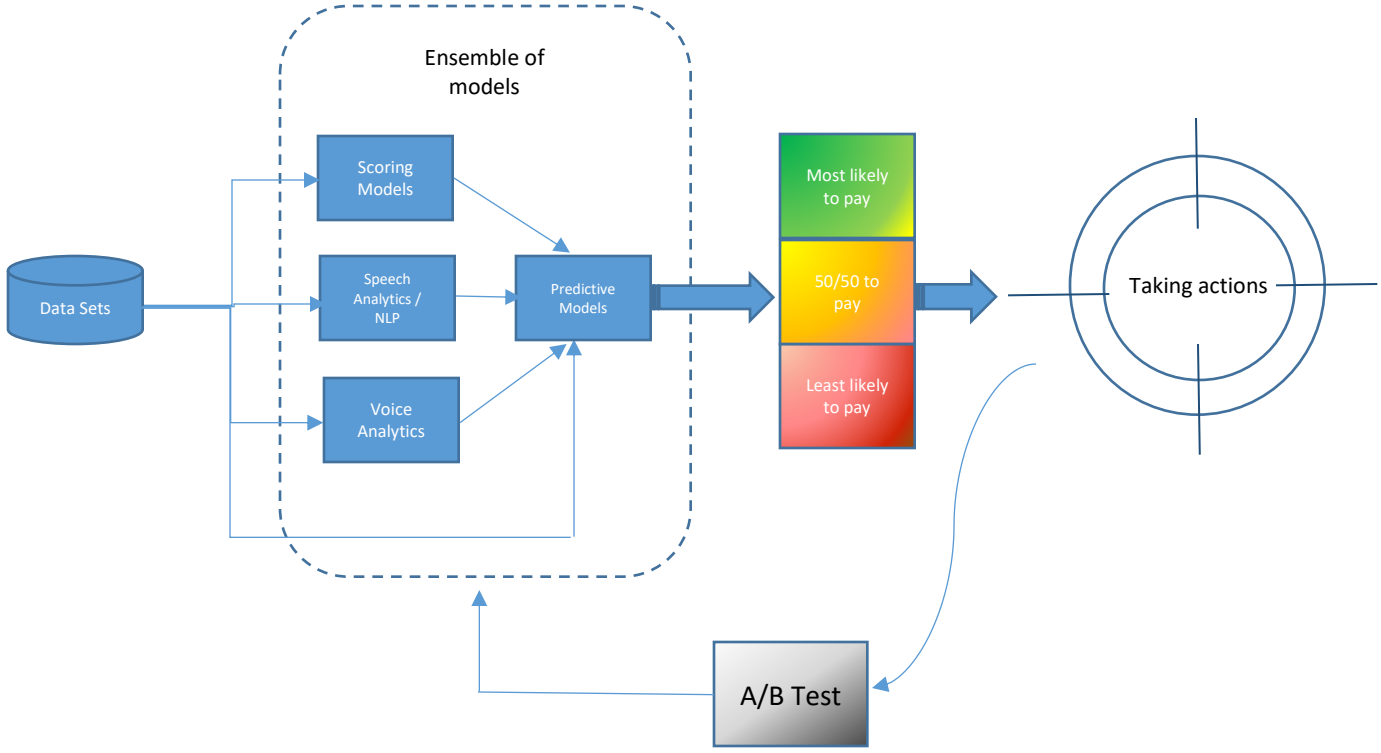
3. Machine Learning

I would need to combine multiple models:

- Scoring models taking into account a client credit score, demographics, history of debts/payments etc.
- NLP and Voice Analytics models to get the content and the context of conversations.
- Predictive models explaining:
    - Correlation of success or failure outcomes with patterns in the agent and debtor conversation
    - Criteria of the calls to be successful
    - The final score showing how likely the debtor to pay

The output of this step will produce actionable insights to support decision making on a particular debtor. The purpose is to maximize the amount collected from clients who are more likely to pay and minimize the cost of debts collection from those who cannot pay.

4. Run A/B test between the predictive suggestions and regular call practices to prove the obtained results or make necessary improvements in the developed models.



Additional Thoughts on Further Steps

- I would be also interested in investigating how to leverage social media (Facebook, Twitter) data to trace debtor. Using data on e-mails, SMS may be useful as well.

- Productionizing the ML Pipeline. It may be a standalone program making ad-hoc predictions. Or it may be incorporated in the existing infrastructure (enterprise software, data lake). A huge topic to think about how to deploy the model and a lot of exciting work (probably with Docker, REST API) on the actual implementation.