

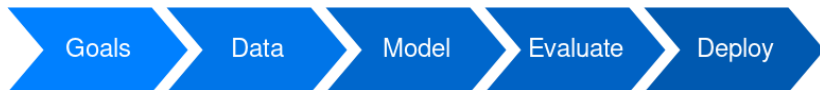
# Proposed Analytic Framework

Tuomo Kareoja

Alert! Analytics

November 3, 2019

# Analytic Process



Why do we want to do this?

What do we need from this?

Why is current solution inadequate?

What data is available?

Is there enough data?

Is the data quality good enough?

How will the data be handled?

Which techniques to apply to build the model?

How many different techniques should we apply?

Is the model accurate enough to meet our needs?

Does the model perform better than "the obvious guess" and any techniques being used currently?

Is the model interpretable?

When should model predictions be manually overruled?

How is the model to be handed off to "production"?

How often, and under which circumstances, should the model be revised?

# Goals

- ▶ Why do we want to do this?

# Goals

- ▶ Why do we want to do this?
  - ▶ Increasing number of customers who default

# Goals

- ▶ Why do we want to do this?
  - ▶ Increasing number of customers who default
  - ▶ Risk of losing customers to other credit rating agencies if our predictions don't get better

# Goals

- ▶ Why do we want to do this?
  - ▶ Increasing number of customers who default
  - ▶ Risk of losing customers to other credit rating agencies if our predictions don't get better
- ▶ What do we need from this?

# Goals

- ▶ Why do we want to do this?
  - ▶ Increasing number of customers who default
  - ▶ Risk of losing customers to other credit rating agencies if our predictions don't get better
- ▶ What do we need from this?
  - ▶ Better predictive model, so that less people who would default on their loans don't get loans or get only smaller loans

# Goals

- ▶ Why do we want to do this?
  - ▶ Increasing number of customers who default
  - ▶ Risk of losing customers to other credit rating agencies if our predictions don't get better
- ▶ What do we need from this?
  - ▶ Better predictive model, so that less people who would default on their loans don't get loans or get only smaller loans
- ▶ Why is the current solution inadequate?



# Goals

- ▶ Why do we want to do this?
  - ▶ Increasing number of customers who default
  - ▶ Risk of losing customers to other credit rating agencies if our predictions don't get better
- ▶ What do we need from this?
  - ▶ Better predictive model, so that less people who would default on their loans don't get loans or get only smaller loans
- ▶ Why is the current solution inadequate?
  - ▶ Too many will-be-defaulters get loans or the get loans that are too big

# Goals

- ▶ Why do we want to do this?
  - ▶ Increasing number of customers who default
  - ▶ Risk of losing customers to other credit rating agencies if our predictions don't get better
- ▶ What do we need from this?
  - ▶ Better predictive model, so that less people who would default on their loans don't get loans or get only smaller loans
- ▶ Why is the current solution inadequate?
  - ▶ Too many will-be-defaulters get loans or the get loans that are too big
  - ▶ QUESTION? Why don't make the current model more stringent?

# Data

- ▶ What data is available?

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns
  - ▶ Data is from 2005. Has there been no significant changes since then?



# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns
  - ▶ Data is from 2005. Has there been no significant changes since then?
- ▶ Is there enough data?

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns
  - ▶ Data is from 2005. Has there been no significant changes since then?
- ▶ Is there enough data?
  - ▶ about 20 % of customers default in the data. Is this the true rate for all customers?

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns
  - ▶ Data is from 2005. Has there been no significant changes since then?
- ▶ Is there enough data?
  - ▶ about 20 % of customers default in the data. Is this the true rate for all customers?
  - ▶ More data would mean more precise models. Even 10X current data size would be easily handled

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns
  - ▶ Data is from 2005. Has there been no significant changes since then?
- ▶ Is there enough data?
  - ▶ about 20 % of customers default in the data. Is this the true rate for all customers?
  - ▶ More data would mean more precise models. Even 10X current data size would be easily handled
- ▶ How will the data be handled?

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns
  - ▶ Data is from 2005. Has there been no significant changes since then?
- ▶ Is there enough data?
  - ▶ about 20 % of customers default in the data. Is this the true rate for all customers?
  - ▶ More data would mean more precise models. Even 10X current data size would be easily handled
- ▶ How will the data be handled?
  - ▶ Locally on one computer

# Data

- ▶ What data is available?
  - ▶ Information about 30000 customers payment default status with demographic information and 6 month of information about previous loan amounts and payments
- ▶ Is the data quality good enough?
  - ▶ Almost no missing data
  - ▶ Dataset information given does not match the data on all columns
  - ▶ Data is from 2005. Has there been no significant changes since then?
- ▶ Is there enough data?
  - ▶ about 20 % of customers default in the data. Is this the true rate for all customers?
  - ▶ More data would mean more precise models. Even 10X current data size would be easily handled
- ▶ How will the data be handled?
  - ▶ Locally on one computer
  - ▶ Erased at the end of the project

The End

Questions?