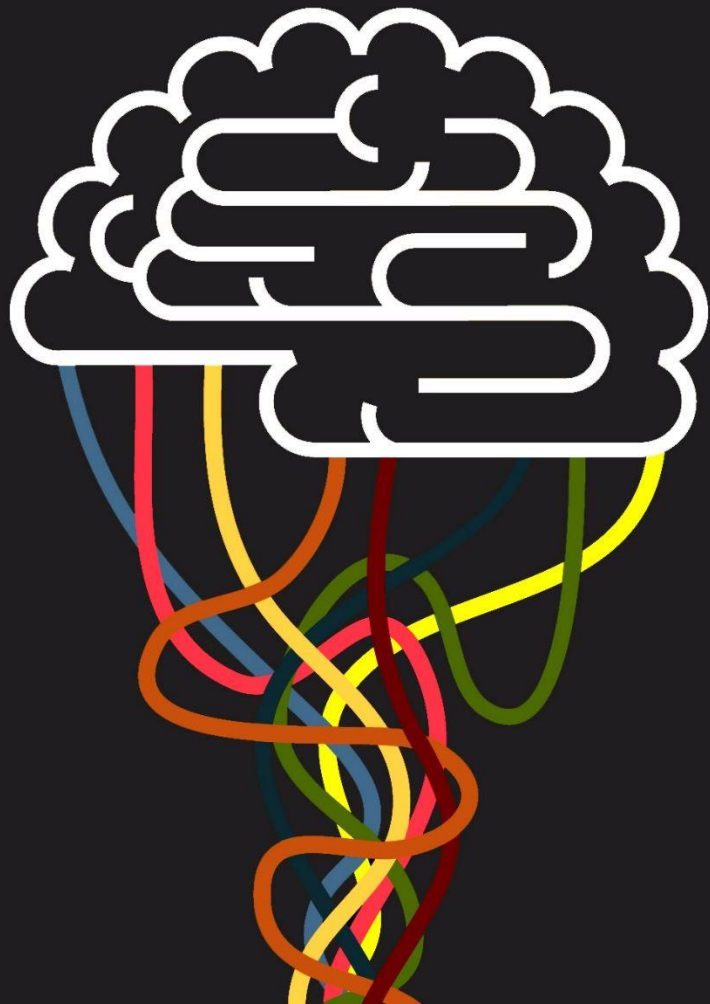




Interactive Development in Jupyter Notebooks

ML Intro Training

July, 2019



Training

Topic:

Interactive development in Jupyter Notebooks

Why:

Most of Datathon cases would benefit from elements of AI/ML*

*learning from data

Objectives:

Get hands-on experience with explorative data analysis, ML model training, evaluation & deployment

Take-away points:

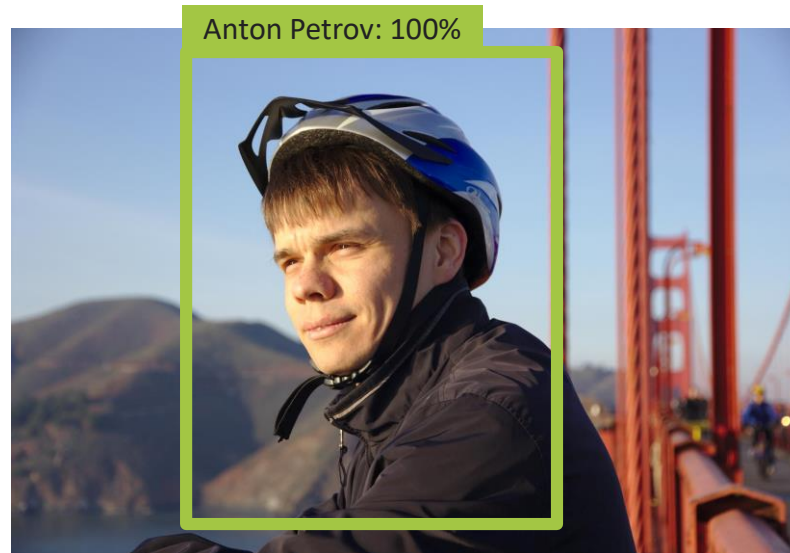
AI/ML is hard, but prototyping is quick

About Me

Anton Petrov

Lead Data Scientist, EPAM

- Prototyping ML solutions
- Consulting on ML architecture
- Team mentoring and coaching



R&D

Product Company

Finance

Consulting



Agenda

Environment



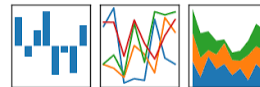
20 mins



30 mins

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Microsoft
LightGBM

mlflow



10 mins



40 mins



Business Case

Background

People talent is the most valuable resource we have.

Reduction of the team size increases time to market of new products and leads to revenue shrinkage, while higher employee turnover rate results in increased costs due to recruiting and training, all which negatively affects the bottom line.

HR team does their best and is constantly looking for new insights, ideas and approaches.

Challenge

Leverage historical data to identify employees likely to leave the company before they act, reveal the driving factors, provide data-driven suggestions and insights to the leadership team.

Business Case

Background

People talent is the most valuable resource we have.

Reduction of the team size increases time to market of new products and leads to revenue shrinkage, while higher employee turnover rate results in increased costs due to recruiting and training, all which negatively affects the bottom line.

HR team does their best and is constantly looking for new insights, ideas and approaches.

attrition → less revenue, more costs → reduction of profit



Challenge

Leverage historical data to identify employees likely to leave the company before they act, reveal the driving factors, provide data-driven suggestions and insights to the leadership team.



identify employees

driving factors

insights

suggestions

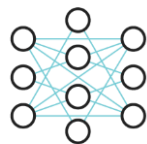


ML-first Approach



Data

fit



ML Model

predict

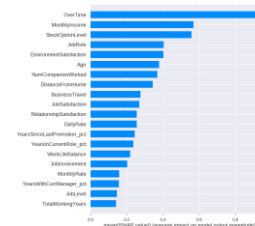


identify employees

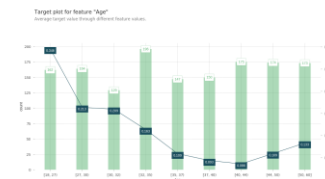
EmployeeNumber	
167	0.996485
1053	0.908141
1928	0.897252
1244	0.877069
2021	0.870730

groups / clusters of employees

driving factors



insights



suggestions

