

How to become a Kaggle Competitions Expert.

Data Science & Engineering Club

24th March 2018



Hi all!

I am Roman Golovnya, the organiser of this meetup group and a part-time kaggler.

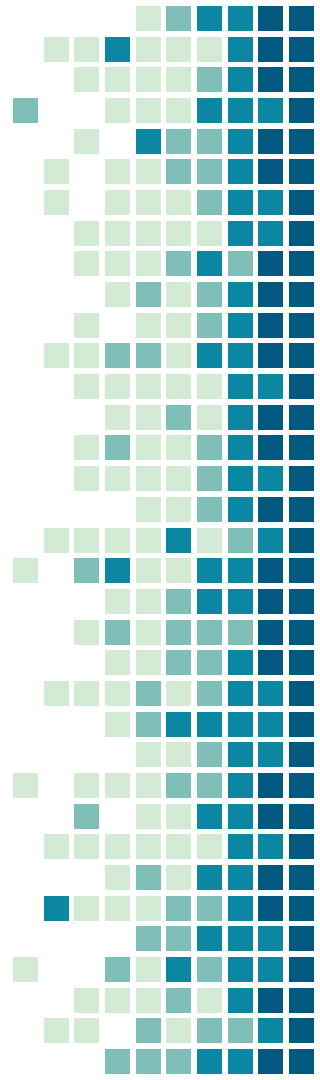
I work as a Data Engineer at The Eclair Group - the fast growing automation and AI start-up.

roman.golovnya@gmail.com

<https://www.linkedin.com/in/roman-golovnya-11a6b614>

<https://www.kaggle.com/rgolovnya>

<https://github.com/roman79/DSEC>



Kaggle is the best known machine learning and predictive modelling competition platform.

[Kaggle](#) is a great place to learn from machine learning gurus and data science evangelists. More than 1 million members.

<https://www.kaggle.com/rankings>

Competitions

Kernels

Discussion

Learn more about rankings ›



108

Grandmasters



1,002

Masters



3,193

Experts



47,971

Contributors



28,171

Novices

Rank	Tier	User	Medals	Points
728		 <div>Roman</div> <div>joined 2 years ago</div>	<div>  0  1  5 </div>	9,086

Kaggle statuses: Novice – Contributor are very easy to get!



Expert

You've completed a significant body of work on Kaggle in one or more categories of expertise. Once you've reached the expert tier for a category, you will be entered into the site wide Kaggle Ranking for that category.

Competitions

✓ 2 bronze medals

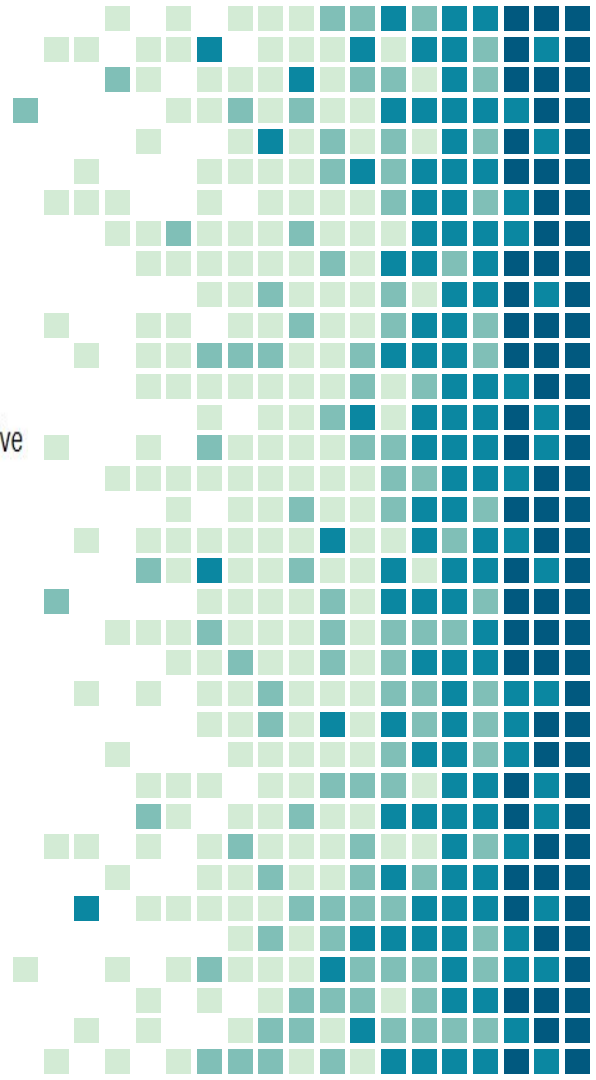
Kernels

□ 5 bronze medals

Discussions

□ 50 bronze medals

<https://www.kaggle.com/progression>



Kaggle Toolsets

- The most popular programming/scripting languages are Python and R.
- Pandas, Scikit-Learn, XGboost, LightGBM, Tensorflow, Keras,
- Docker, Linux,
- Cloud resources.



Kaggle has many learning resources.

<http://blog.kaggle.com/category/tutorials/>

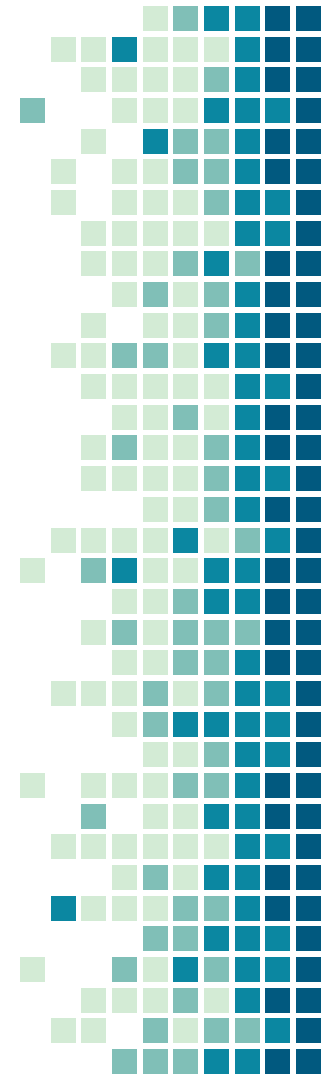
<https://www.kaggle.com/learn/machine-learning>

Kaggle hands-on Data Cleaning training (free).

5 -Day Challenge: Data Cleaning

March 26 - March 30, 2018

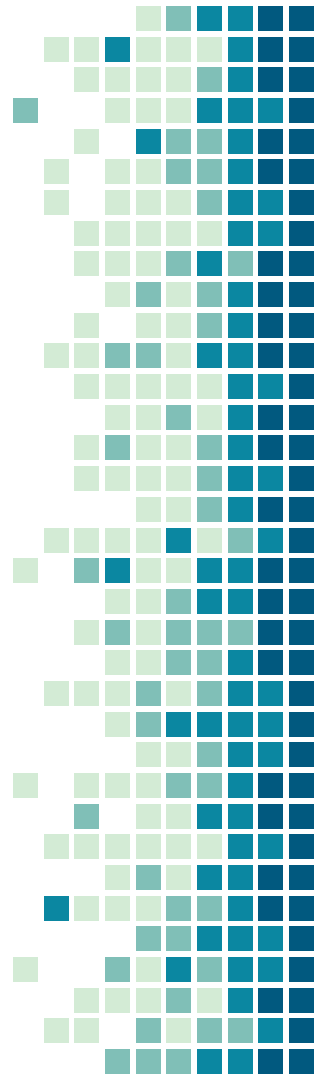
- Day 1: Handling missing values
- Day 2: Data scaling and normalization
- Day 3: Cleaning and parsing dates
- Day 4: Character encoding errors (no more messed up text fields!)
- Day 5: Fixing inconsistent data entry & spelling errors



Kaggle competition approach - Kaggle forum helps a lot!

- Understand the completion question, metrics and what is required to be submitted
- Load data
- Data Preprocessing
- Exploratory Data Analysis
- Create a Single Model
- Evaluate the model
- Submit results - benchmark
- Feature Engineering
- Create Better Models
- Cross - Validation
- Evaluation of models
- Ensembles

Next



To climb higher you will need to develop your own expertise around feature engineering, algorithm tuning, pipelining, deep learning, ensembles.

Excellent material on stacking or ensembling [Kaggle Ensembling Guide](https://mlwave.com/kaggle-ensembling-guide/).<https://mlwave.com/kaggle-ensembling-guide/>

Conclusion

- Find the competition with data which you like
- Team-up with others
- Learn from top Kaggle winners, discussion forums, blogs
- Be organised
- Prepare your reusable packages or templates
-



THANKS!

Any questions?

You can contact me at:

roman.golovnya@gmail.com