MACHINE LEARNING WITH PYTHON FOR SPACE WEATHER APPLICATIONS

by ITU Upper Atmosphere and Space Weather Laboratory

Lecture 5: Hands on Space Weather Application

Instructor: Dogacan Su Ozturk

Contact: dsozturk@alaska.edu

11-12 May 2022 at Upper Atmosphere and Space Weather Laboratory

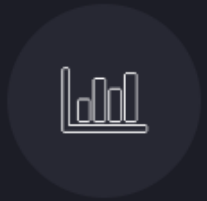
The 8 ML Principles from the Institute for Ethical AI & Machine Learning

1. Human augmentation

I commit to assess the impact of incorrect predictions and, when reasonable, design systems with human-in-the-loop review processes

2. Bias evaluation

I commit to continuously develop processes that allow me to understand, document and monitor bias in development and production.



3. Explainability by justification

I commit to develop tools and processes to continuously improve transparency and explainability of machine learning systems where reasonable.

4. Reproducible operations

I commit to develop the infrastructure required to enable for a reasonable level of reproducibility across the operations of ML systems.

5. Displacement strategy

I commit to identify and document relevant information so that business change processes can be developed to mitigate the impact

6. Practical accuracy

I commit to develop processes to ensure my accuracy and cost metric functions are aligned to the domain-specific applications.

7. Trust by privacy

I commit to build and communicate processes that protect and handle data with stakeholders that may interact with the system directly

8. Data risk awareness

I commit to develop and improve reasonable processes and infrastructure to ensure data and model security are being taken into

Resources for an Ethical ML Framework

🔍 Explaining predictions & models	🪶 Privacy preserving ML	📜 Model & data versioning
🏁 Model Training Orchestration	💪 Model Serving and Monitoring	🤖 Neural Architecture Search
📓 Reproducible Notebooks	📊 Visualisation frameworks	🔤 Industry-strength NLP
🧵 Data pipelines & ETL	🏷️ Data Labelling	📆 Metadata Management
📡 Functions as a service	🗺️ Computation distribution	📥 Model serialisation
🧮 Optimized computation frameworks	🌠 Data Stream Processing	🔴 Outlier and Anomaly Detection
🌀 Feature engineering	🎁 Feature Stores	⚔️ Adversarial Robustness
💰 Commercial Platforms	💾 Data Storage Optimization	

https://github.com/EthicalML/awesome-production-machine-learning#model-and-data-versioning

8 Main Steps for Machine Learning Process

1. Frame the problem and look at the big picture.
2. Get the data.
3. Perform exploratory analysis.
4. Prepare the data for the ML applications.
5. Explore different models and shortlist the best ones.
6. Fine-tune the models and combine them into a solution.
7. Present your solution.
8. Launch, monitor, and maintain your system.

1. Frame the problem and look at the big picture.

1. Define the objective in business/research terms.
2. Determine how your solutions will be used.
3. Identify the current solutions/workarounds (if any).
4. Determine how to frame the problem (supervised/unsuperv
5. Determine performance metrics.
6. Check if the performance metric is aligned with the busines
7. Determine the minimum performance necessary to reach th
8. Determine if the experience and tools are transferrable.
9. Check if human expertise is available.
10. Identify how the solution would look like manually.
11. List all assumptions.
12. Verify the assumptions.

Image Credit: Double Take by Susan Hood and Jay Fleck

2. Get the data.

1. List the data you need and how much of it you need.
2. Find and document where you can get the data.
3. Check how much space the data will take.
4. Check legal obligations (and ethical!). Acquire authorization if necessary.
5. Create a workspace.
6. Get the data.
7. Convert the data to a format you can easily manipulate.
8. Ensure sensitive information is protected.
9. Check the size of the data.
10. Sample a test set, put it aside, and never look at it until its time.

Illustration by Andrew Krasovitckii

3. Perform exploratory analysis.

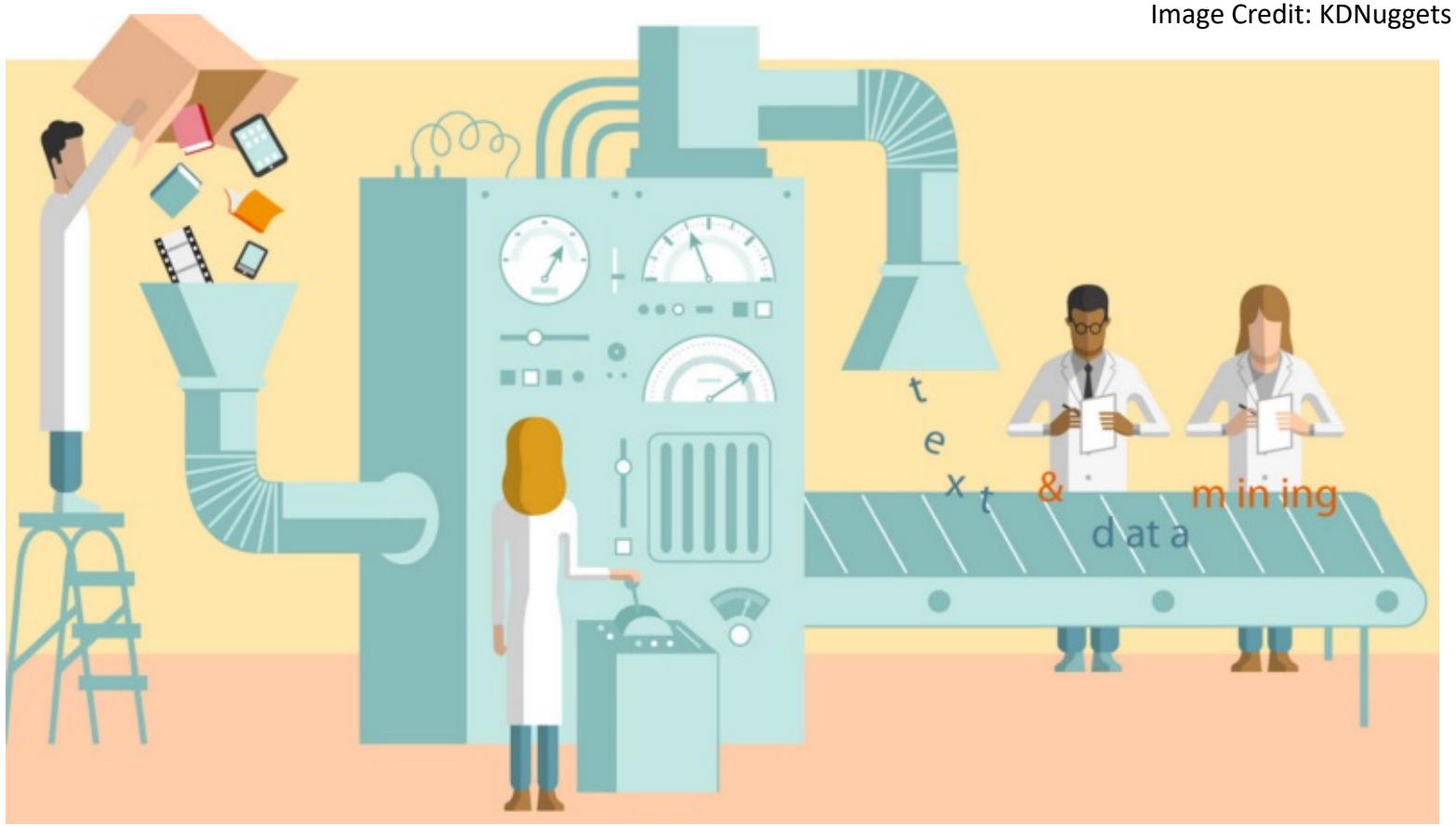
1. Create a copy of data for exploration.
2. Create a Jupyter notebook to keep a record of your exploration.
3. Study each attribute and characteristics:
    a. Name
    b. Type
    c. % of missing values
    d. Noisiness
    e. Usefulness
    f. Distribution
4. Identify feature and target for supervised tasks.
5. Visualize the data.
6. Study the correlations between attributes.
7. Study how you would solve the problem manually.
8. Identify necessary transformations.
9. Document what you have learnt.


Image Credit: Visme

4. Prepare the data for the ML applications.

1. Clean the data.
   a. Remove outliers.
   b. Handle NaNs.
2. Select features.
3. Engineer features if necessary.
4. Scale features (omit if using pipelines).

Image Credit: KDNuggets



5. Explore different models and shortlist the best ones

1. Train using many easy models from different categories using standard parameter grids.
2. Measure and compare their performance.
3. Analyze most significant variables for each algorithm.
4. Analyze types of error models make.
5. Shortlist some of the most promising models, making sure to pick a variety that make different types of errors.

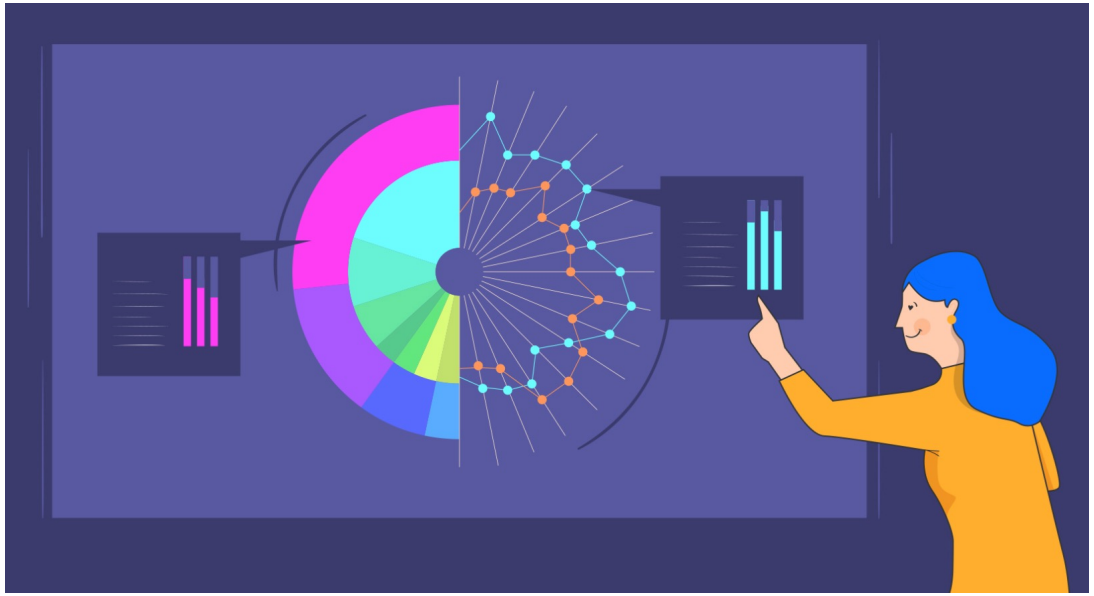


Image Credit: Visme

6. Fine-tune the models.

1. Fine-tune hyperparameters using cross validation.

2. Try ensemble methods.

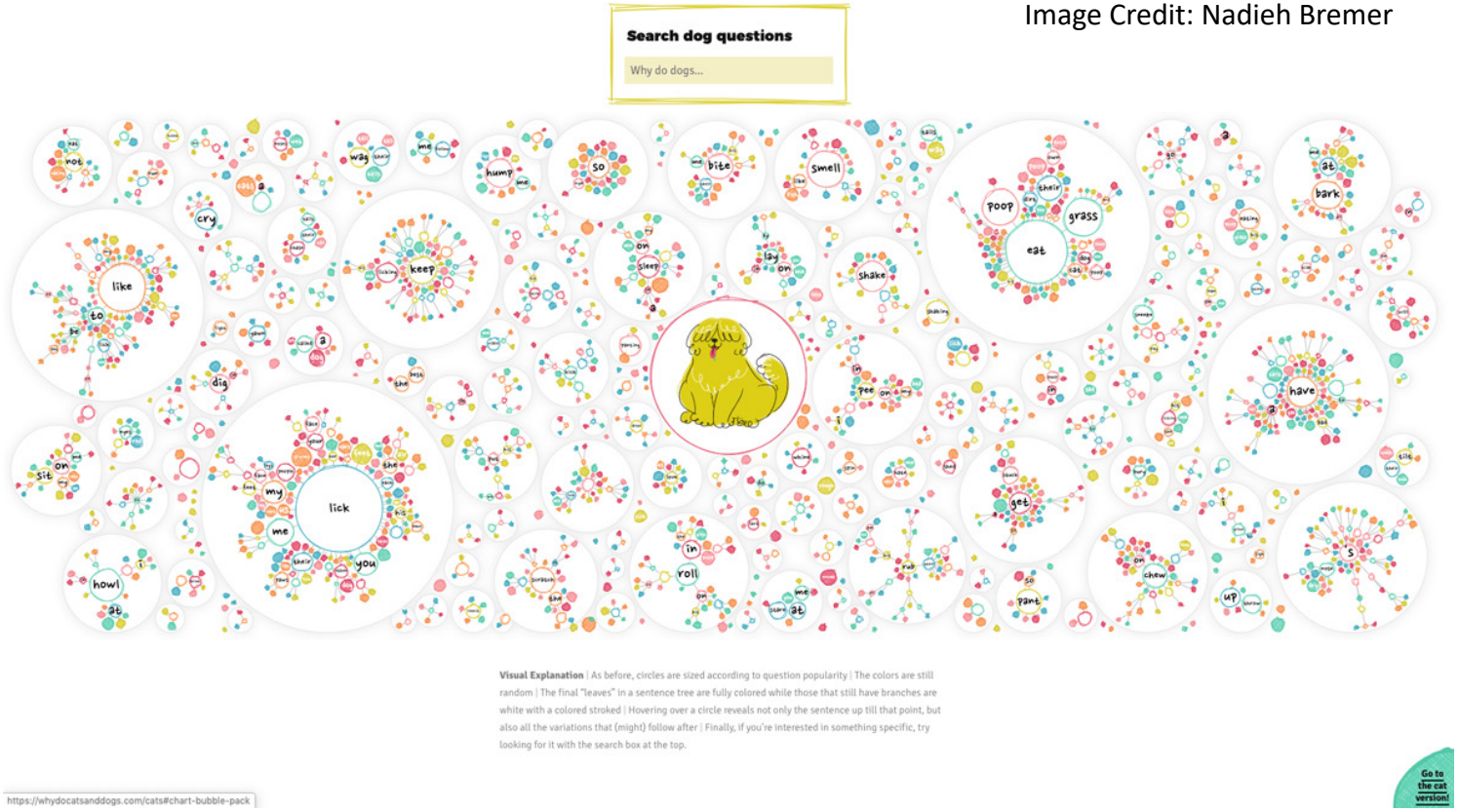
3. Once you settle on a model, measure its performance on the test set.

Man Tuning a Violin by Frans van Mieris

7. Present your solution.

1. Document your activities.
2. Create a presentation.
3. Explain why your solution achieves business/research objectives.
4. Present interesting findings from your analyses.
5. Ensure key findings are communicated, beautiful visualizations are handy!

Image Credit: Nadieh Bremer



8. Launch, monitor, and maintain your system.

1. Make sure everything is automated.
2. Get your solution ready for production.
3. Write monitoring code to check your system's performance.
4. Retain your models on a regular basis on fresh data.

Image Credit: Freepik

Image Credit: John L. Bates and Associates