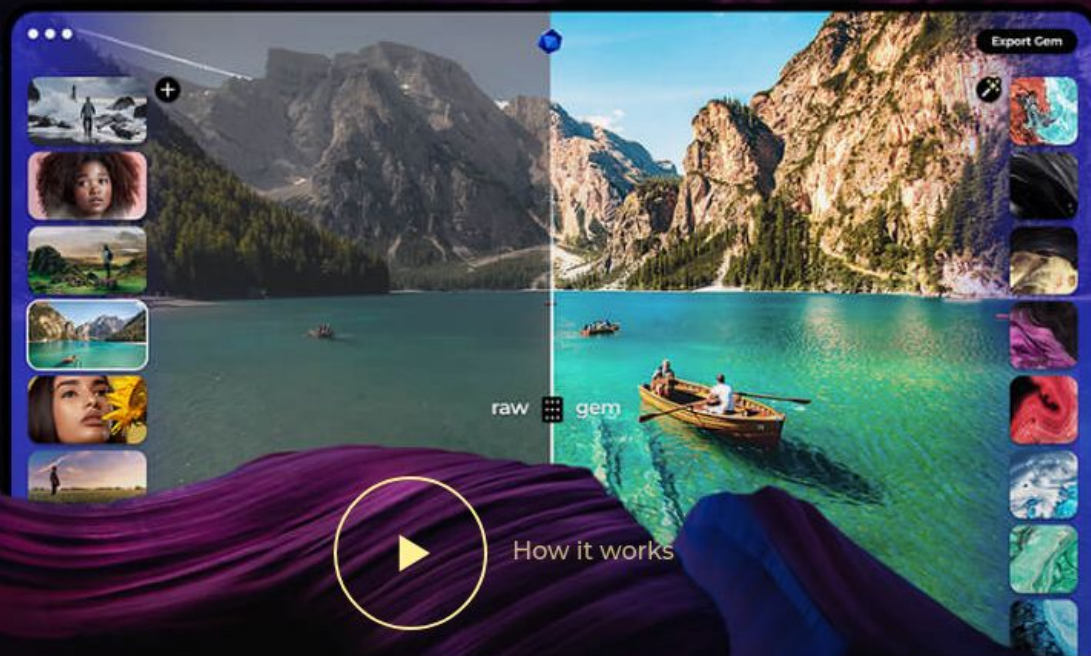


The lifecycle of creating a real AI-based product

Kateryna Stetsiuk
Head of R&D at ARVI



Gemia.me



Gemia is easy to use desktop Photo Editor for automatically photo enhancement.

Just upload photos and get an enhanced result in a few seconds.

Buy Now

Know more

Available on the
App Store

Gemia.me



Gemia.me

Before

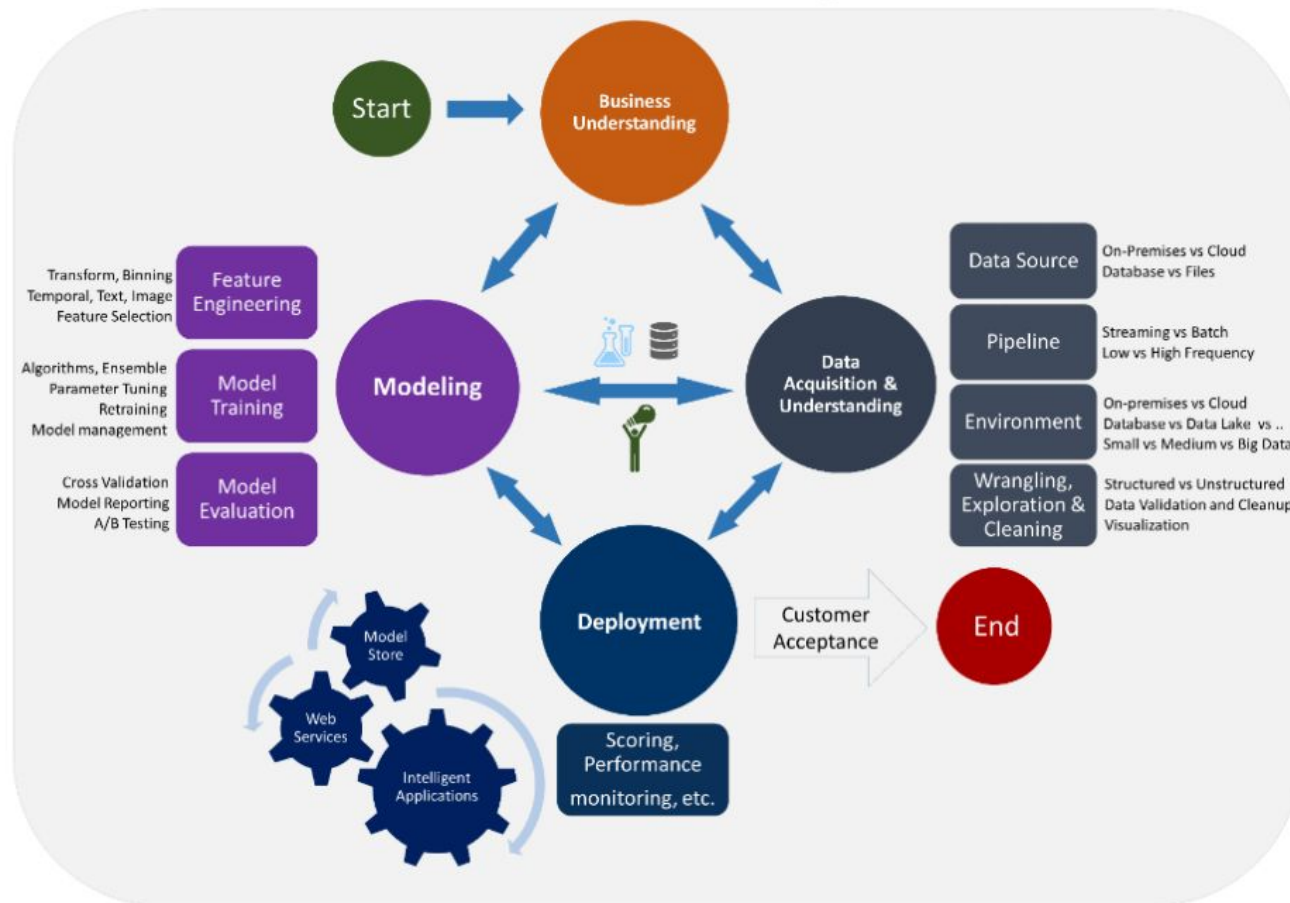


After

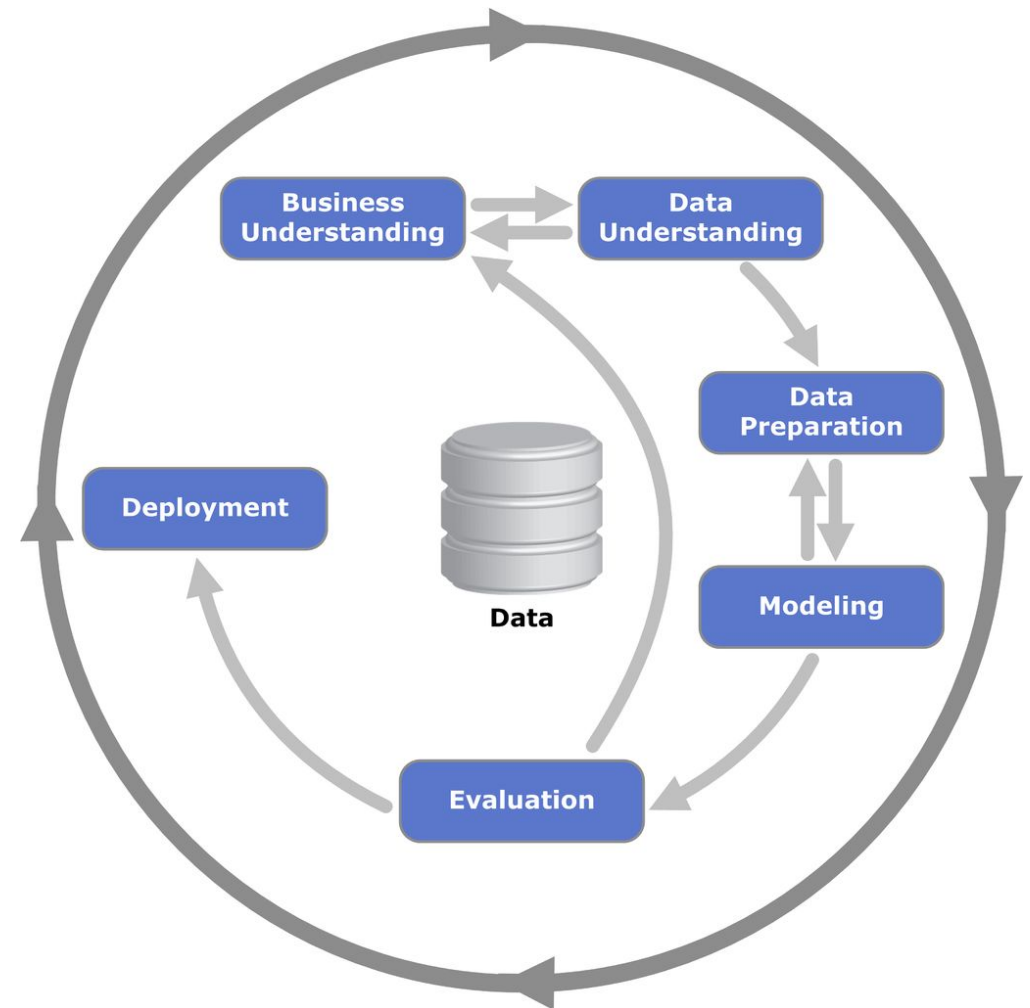


AI Product Development Methodology

TDSP



CRISP-DM



Business Understanding

- **Define objectives:** identify business problems and define the goals.
- **Identify data sources:** find the relevant data that helps you to answer the questions that define the objectives of the project.
- **Assessing current situation:** check if there's enough hardware, data, expertise.
- **Define the success metrics** that will be used to assess the success of the project
- **Develop a high-level milestone plan** that you iterate on as you discover more information.

Risk Evaluation

Evaluate Task = Evaluate Risks

- Lack of data
- Lack of knowledge
- Lack of resources
- Time to train models
-



Gemia Case. Business Understanding

- Communicate with photographers about problems
- Define goals and satisfying results

Data Understanding & Preparation



Data Understanding & Preparation

Understanding

- data collection
- data description
- data exploration
- data quality

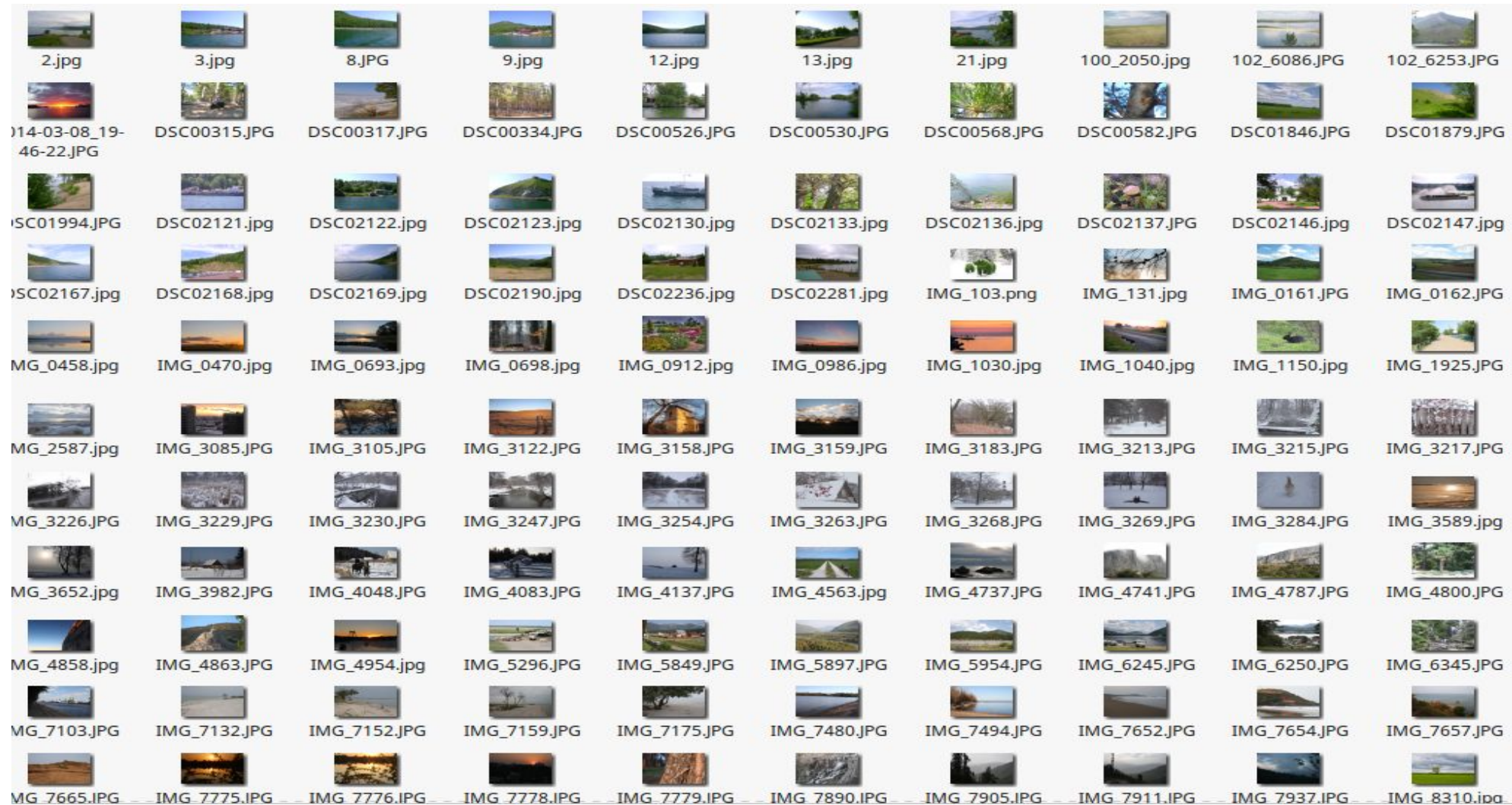
Preparation

- data selection
- data cleaning
- constructing new data
- data integration

Gemia Case. Data Understanding & Preparation

- Gathering data from photographers - **10 000 pairs of photos**
- Understand existing data problems and decide the solution
- Data cleaning
- Prepare data for a training

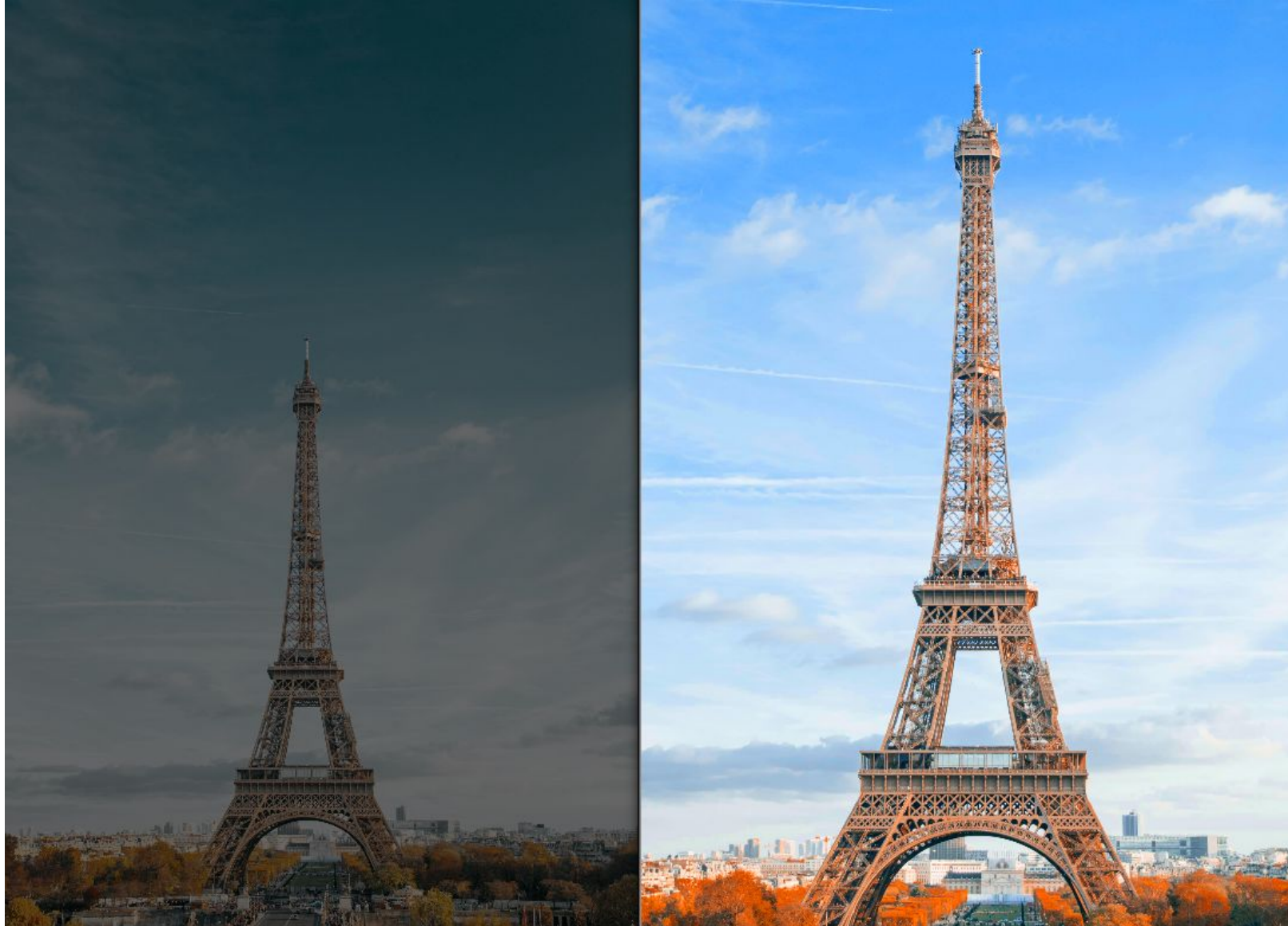
Problem#1 – Unpaired photos



Problem#2 - Different sizes



Problem#3 – Different cut



Modeling Stage

- Selecting the modeling technique
- Generating test design
- Building the models
- Assessing the models



Hypothesis & Modeling

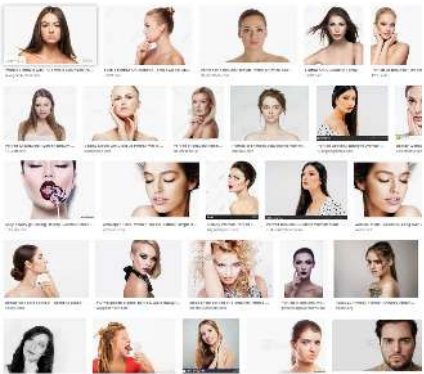
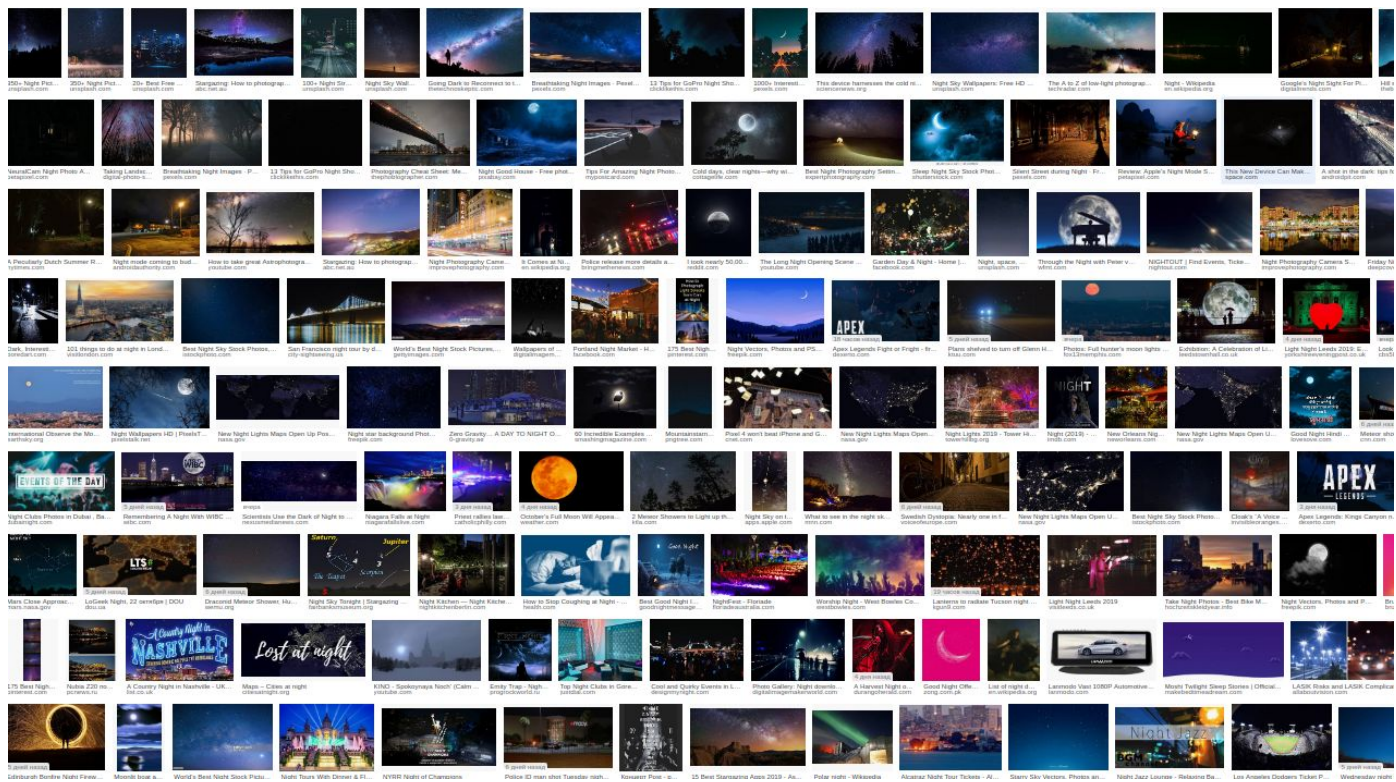
- Create hypothesis and work with them **in sprints**.
- Stay nimble and try many **parallel (isolated) ideas** on early stages
- Find **SoTA** model for your problem's domain (if available) and reproduce results, then apply to your dataset as a second baseline
- Perform model-specific optimizations (ie. hyperparameter tuning)

Gemia Case. Modeling

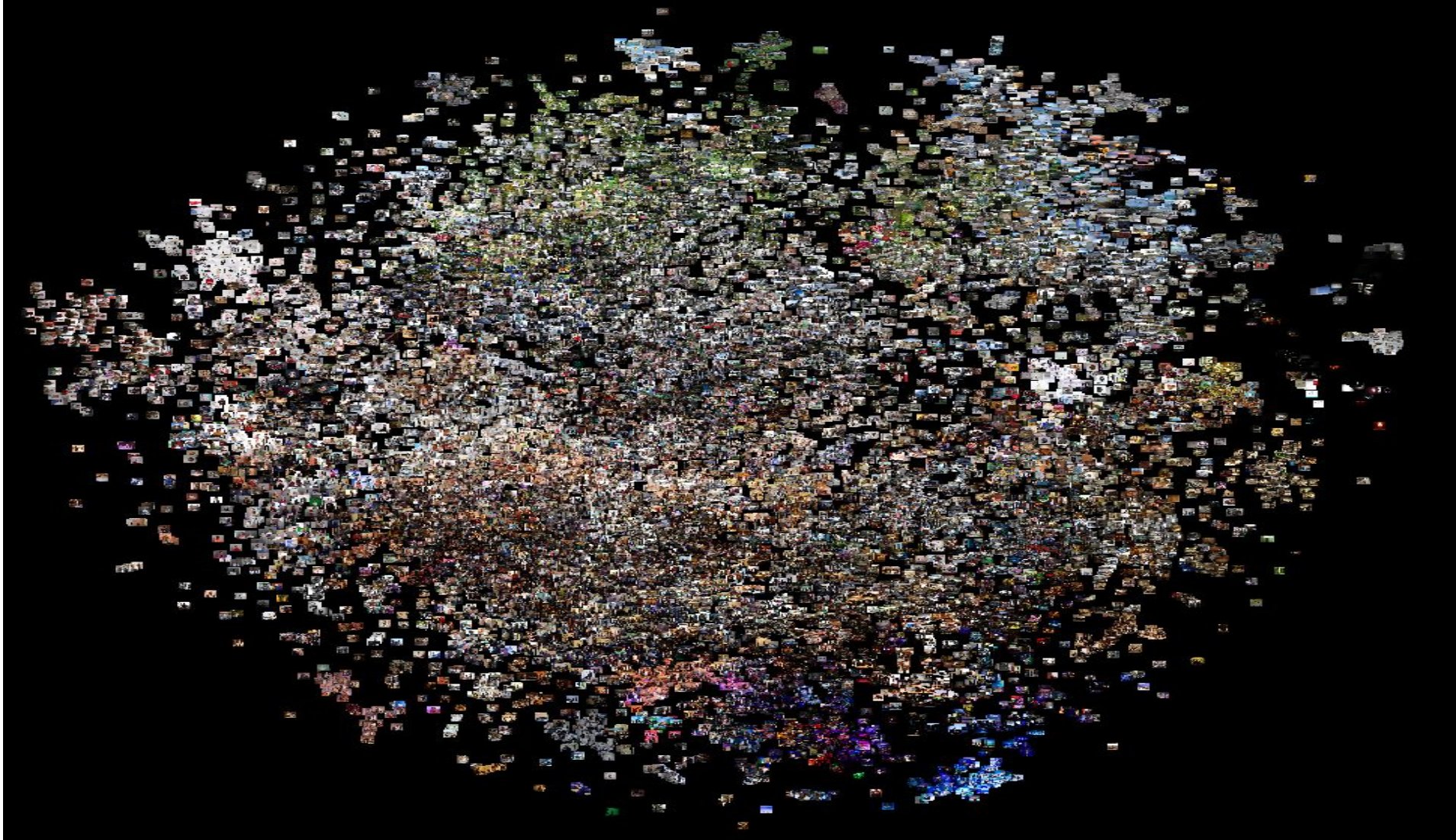
- Choose different approaches (classical computer vision, neural networks)
- The easiest (fastest) first
- Experiments with datasets



Problem#1 – Unbalanced datasets



Problem#1 – Unbalanced datasets



Problem#1 – Unbalanced datasets



Problem#2 – Incorrect Photos



A lot of manual work ...



Problem#3 - Similar but not the same photos



Problem#4 – Different evaluation from photographer and network





Problem#5 – Processing Time





Evaluation Stage

- Evaluating the results (with domain experts)
- Review the process- figure out all the problems
- Determining the next steps - how to improve

Gemia Case. Evaluation

- Create a demo to test
- Gather different photographers' feedback
- Improve models during testing



Deployment Stage

- Planning Deployment
- Planning Monitoring
- Final Report

Planning Deployment

To deploy models, you expose them with an open **API interface**. The interface makes the model easily consumed from various applications, such as:

- Online websites
- Dashboards
- Line-of-business applications
- Back-end applications

Gemia Case. Deployment

- Create an API for integration
- Create online version of the application
- Run models on CPU
- Create Mac OS application
- Improve models till now



Problem#1 – Run on CPU

Problem#2 – Optimize run time

Common Mistakes in Creating AI Product

- Business and Data Understanding Mistakes
- Data preparation and modeling Mistakes
- Production Mistakes



Business and Data Understanding Mistakes

- The problem is not validated by the customer.
- Potential solution is not clear to the customer and he does not buy it
- Do not listen to experts
- Not enough data analysis before further steps

Data Preparation and Modeling Mistakes

- Upload incorrect data
- Forgot about the area; testing experiment
- Cannot reproduce the results

Production Mistakes

- On the historical data we received “outstanding” results with high model quality
- Missing model quality monitoring

Thank You!

Contacts:



<https://www.facebook.com/katiastetsiuk>



<https://www.linkedin.com/in/katernastetsiuk/>



katiastet@gmail.com