Razvijanje rešitev umetne inteligence v sodelovanju s SLAIF

Uroš Lotrič, UL FRI

Cilji

- Prelivanje akademskih rešitev v prakso
- Množica raznovrstnih aplikacij
- Uporaba nacionalnih in evropskih superračunalniških in oblačnih virov

- Opolnomočenje MSP za pripravljanje in ponujanje rešitev UI na trgu
- Povezovanje uporabnikov in ponudnikov

Zeleni prehod

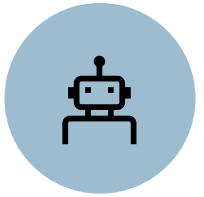
energija, pametna proizvodnja, agroživilstvo

Zdravje in biotehnologija

Diagnostika, osebna medicina, odkrivanje zdravil

Digitalna družba

jezikovne tehnologije, generativni modeli UI, javni sektor, izobraževanje



Znanost

digitalna humanistika, materiali, vede o življenju, medicina, modeliranje

Storitve

Svetovanje pri vključevanju UI v poslovne procese	• povezovanje s kompetentnimi sogovorniki • izbiranje podatkovnih množic, modelov, delovnih tokov • izbiranje ogrodij za delo
Pridobivanje računskih virov	• svetovanje pri izbiri sistema • pomoč pri prijavi na razpise
Podpora pri preoblikovanju ideje v izdelek	• prilagajanje delovnih tokov za ciljne aplikacije • podpora pri pripravi rešitev • načrtovanje potreb za produkcijsko rabo

Katalog rešitev

- Vsebina
  - podatkovne množice
  - modeli
  - delovni tokovi UI
- Navodila za uporabo na nacionalni infrastrukturi
- Učinkovito iskanje: metapodatki, koncept FAIR
- Urednik portala

Primeri storitev – zeleni prehod

Domain-Agnostic Few-Shot Object Counting and Detection

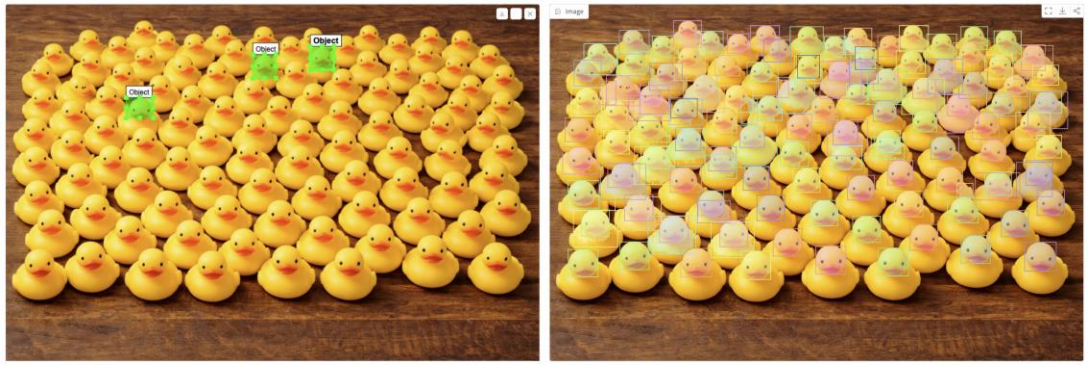
https://huggingface.co/spaces/jerpelhan/GECO2-demo

Target users: Manufacturing companies, industrial SMEs

Responsible: UL FRI

Type of service: model

Availability: Q4/2026



General

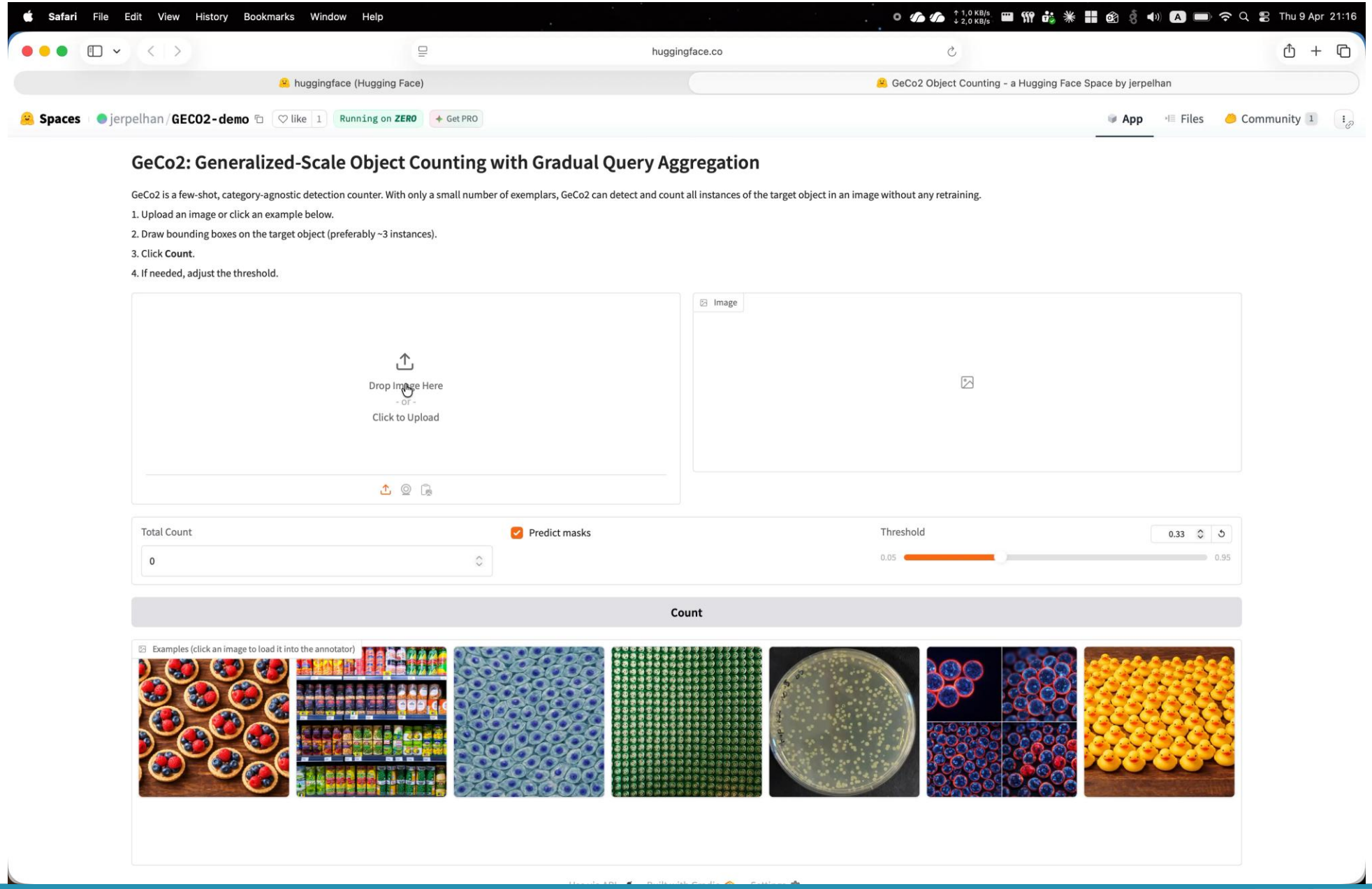
This service enables automated counting and detection of objects in images or videos using only a small number of example annotations. It helps organizations in industry, research, and the public sector quickly gather quantitative insights from visual data, reducing manual effort and enabling rapid deployment across multiple domains. The service provides a real-time inference process to detect and count objects in customer provided images.

Used technologies

The service leverages foundational computer-vision models that generalize from minimal labelled data to accurately detect and count objects across diverse visual domains. By combining feature embedding and clustering techniques the model can identify and count objects of similar shapes, sizes, and appearances, even in complex scenes with cluttered backgrounds. Few-shot learning allows the models to identify similar objects without any extensive retraining or finetuning, making them highly versatile for cross-domain applications. Post-processing and aggregation methods are applied to ensure accurate counting, handle overlapping or partially visible objects, and maintain robust performance under variable lighting or imaging conditions. Code for models used in this service is also available on GitHub:

Expected outcome

The outputs include detection results as bounding boxes and object counts on customer provided images or sequences.



https://huggingface.co/spaces/jerpelhan/GECO2-demo

Primeri storitev – digitalna družba

Adaptation of a Large Language Model to a Specfic Domain

Target users: SMEs, enterprises, public sector, researchers

Responsible: UL FRI

Type of service: pipeline

Availability: Q3/2027

Generativni model za slovenščino
Generative Model for Slovene
GaMS: “

https://arena.cjvt.si/sl

General

This pipeline provides a code and a set of instructions how to adapt a large language model to a specific domain, e.g., specific industry or scientific area. The prerequisites are sufficient amount of domain specific data and domain know-how. It first helps the users in preparing data for instruction following based on the domain data and provides a code and know how how to expand the initial small sample of domain specific instructions into a larger training dataset. Next, it provides the know-how and a code for fine-tuning the model with the created instruction-tuning dataset. Finally, it provides a code for efficient inference using the adapted model. A worked example of adaptation is provided.

Used technology

The pipeline uses Slovene LLM GaMS as well as other state-of-the-art open large language models and libraries that support instruction set expansion, LLM fine tuning with LoRA and QLoRA, and efficient inference with vLLM.

Expected outcome

The outputs include documentation for domain adaptation and code for domain adaptation.

Primeri storitev – zdravje in biotehnologije

FoodyLLM – AI Copilot for Food Nutrition Intelligence

https://huggingface.co/Matej/FoodyLLM

Target users: Nutritionists, food scientists, health data professionals, app developers
Responsible: JSI
Type of service: pipeline

Availability: Q4/2026

General

FoodyLLM is an AI tool designed to analyze and understand food information from text such as recipes, product descriptions, or ingredient lists. It can estimate the nutritional content of recipes (for example fat, carbohydrates, protein, sugar, and salt), predict nutrition labels like the UK FSA traffic-light system, and convert food descriptions into standardized, structured data. This helps companies work with food data more easily, support nutrition analysis, and enable better digital food and health applications.

Used technologies

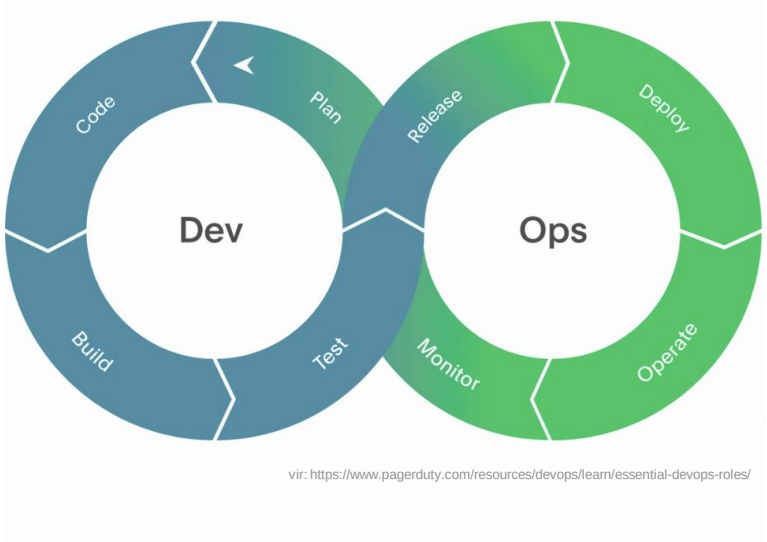
FoodyLLM is built on a transformer-based LLM architecture fine-tuned in a multi-task learning setup to jointly handle macronutrient estimation, FSA traffic-light profiling, and semantic food data mapping. The model leverages food ontologies and nutritional databases during training to ground its predictions.

Expected outcome

Given a recipe's ingredients and quantities (and optionally its title), the model outputs: estimated macronutrient values (fat, carbohydrates, protein, salt, sugar); FSA traffic-light nutritional labels (low/medium/high) per nutrient; and semantically annotated food data aligned with established food ontologies for interoperability with FAIR data repositories and health information systems

Od ideje do izdelka

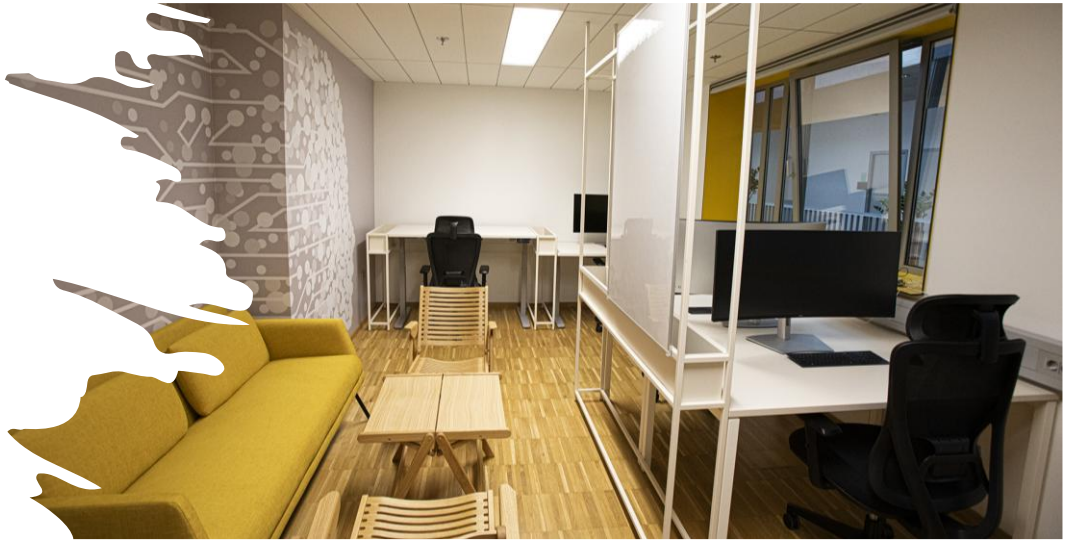
- Ogled izdelkov v katalogu
  - samostjno preizkušanje
  - pomoč pri uporabi modelov na slovenski infrastrukturi
- Povezovanje z raziskovalno skupino
  - svetovanje
  - pomoč pri kuraciji podatkov
- Delo v sodelovalnih prostorih
  - sodelovanje z domenskimi eksperti pri uporabi, prilagajanju, nadgradnji modelov
  - pomoč pri učenju modelov na HPC
- Prehod v predprodukcijsko rabo (PoC)
  - sodelovanje pri načrtovanju produkcijske rabe
  - nameščanje na testno opremo za predprodukcijsko rabo v okolju Kubernets/HPC



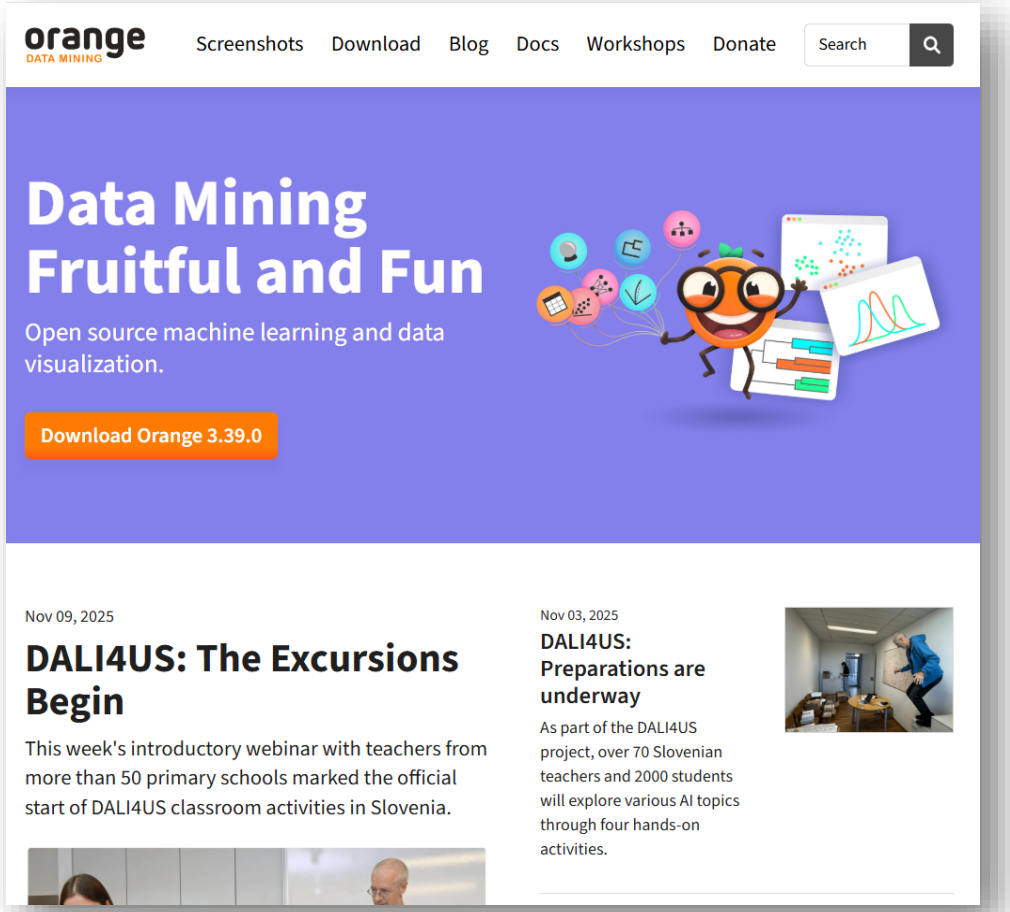
vir: https://www.pagerduty.com/resources/devops/learn/essential-devops-roles/

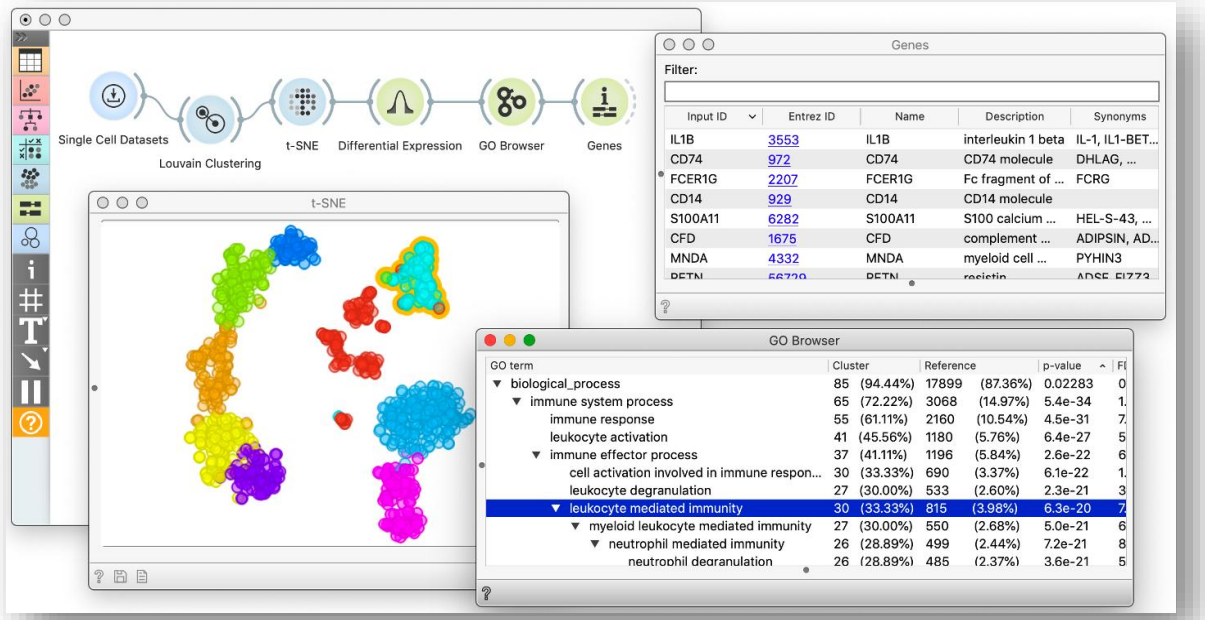
Sodelovalni prostori

- Lokacije: IJS, UL, UM, FIŠ
- UL, pet delovnih mest
  - Nvidia DGX Spark superčip GB10, 128 GB RAM
  - v naprej pripravljeni delovni tokovi za UI https://build.nvidia.com/spark
  - optična povezava na superračunalniško infrastrukturo
  - inferenčni strežnik Nvidia GH200 za testiranje namestitev v oblak
- Dostop za uporabnike SLAIF
  - rezervacijski sistem

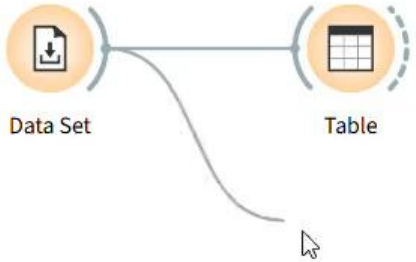


Primeri izdelkov: Orange data mining





orange@biolab.si


Data Set
Table

Primeri izdelkov: govorne tehnologije