Have-a-go series:

Low code/no code ML/DS

Has anyone heard of or used low-code/no-code DS tools before?

nesta

Agenda

- Introduction
- The “So What” and potential for Nesta
- Visual Programming terminology
- Hands-on workflow creation with KNIME
- Examples and use cases
- Further reading

nesta

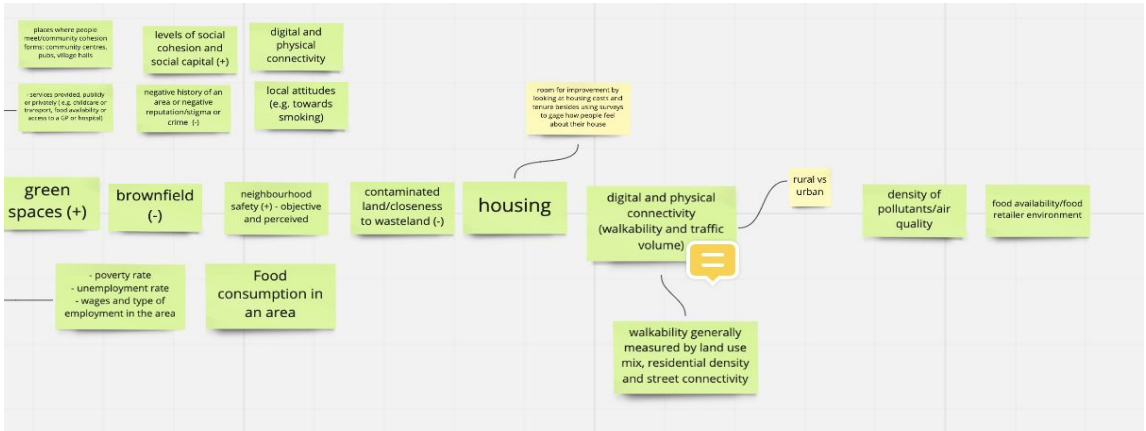
Introduction

nesta

Introduction

- Low code/no code platforms - software that allows you to build software, perform analytics, do data science with little or no coding

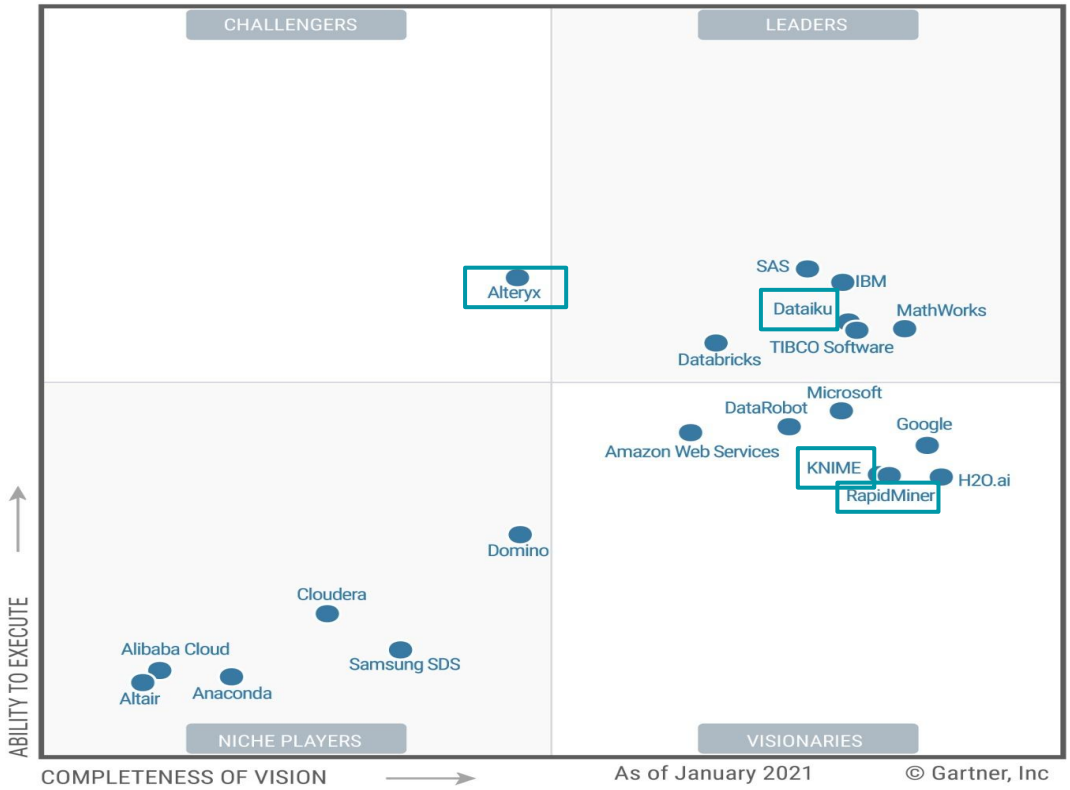
- VPL - visual programming language - uses a graphic interface to carry out steps that you would normally have to program



nesta

Gartner magic quadrant (for internal use only)

Figure 1: Magic Quadrant for Data Science and Machine Learning Platforms



Source: Gartner (March 2021)

nesta

Today we'll be using:

KNIME Analytics Platform

nesta

KNIME Explorer

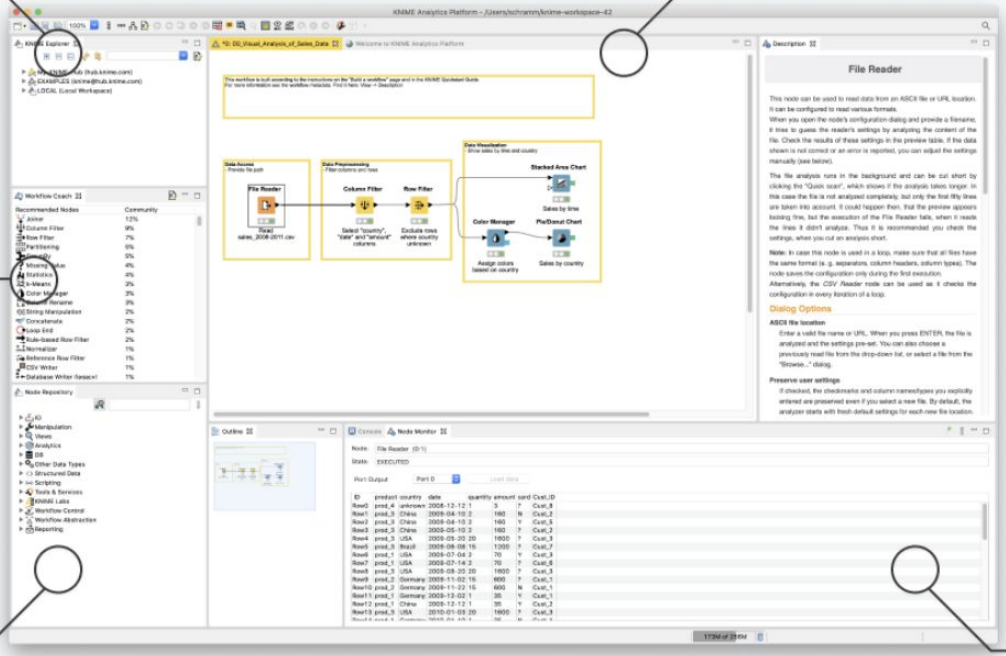
Overview of the available workflows.

Workflow Coach

Lists node recommendations based on the workflows built by the wide community of KNIME users.

Node Repository

All available nodes in KNIME Analytics Platform to build your workflows.



Workflow Editor

Canvas for editing the currently active workflow.

Node Monitor

Shows the current flow variable values or a preview of the output data of the selected node.

The “So What”

nesta

Low-code/no-code platforms lower barriers to entry and development in DS…

What's the potential for DAP and Nesta?

- More transparency and interpretability
- Empowers other teams to perform simple analytics tasks without involving DS (e.g. data wrangling, EDA)
- Easier to share with non-technical experts, less "lost in translation" with developers
- Easy to prototype work early and effectively, and show the reasoning/process
- Flexibility to create websites, models, dashboards etc. without needing to know multiple programming languages, and having a "mental map"
- Write code only when necessary, and customise by integrating Python/R

nesta

But we need to consider the use case.

Downsides

- Need to define scope for usage and security protocols: like giving a child a car to drive
- Deployment is simplified (e.g. creating an API for a model) but requires paid access to a server
- Requires more memory and faster processors
- Does require some programming knowledge for customisation beyond common use cases

nesta

Some visual programming terminology

nesta

KNIME/VP terminology

- Nodes = tasks (e.g. importing data, transforming it, training a model)
- Workflow = project, workflow group = group of nested projects
- A workflow is a collection of nodes, which can only be connected if they have the same data inputs (traffic light system). They have 'ports', or outputs, that you can connect
- A flow variable is an input variable, like an 'x' in an equation

nesta

Hands-on workflow creation

nesta

What we will cover

1. Basic walkthrough of the KNIME Workspace and functionality
2. Example EDA workflow live coding (~10 mins)
3. Simple clustering workflow in groups (~15 mins)
4. A very simple Python integration example
5. Exporting your workflows

nesta

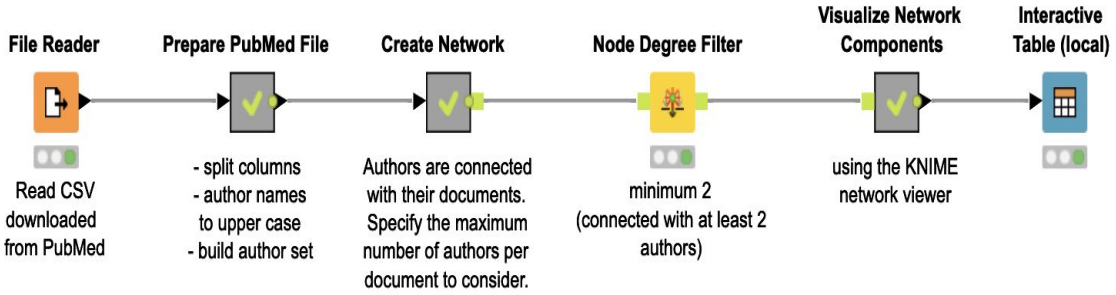
Examples and use cases

nesta

Network analysis

Pubmed network analysis

A workflow that uses network analysis to identify the most important authors of papers which discuss new drugs. A network with authors as nodes and publications as edges is generated.

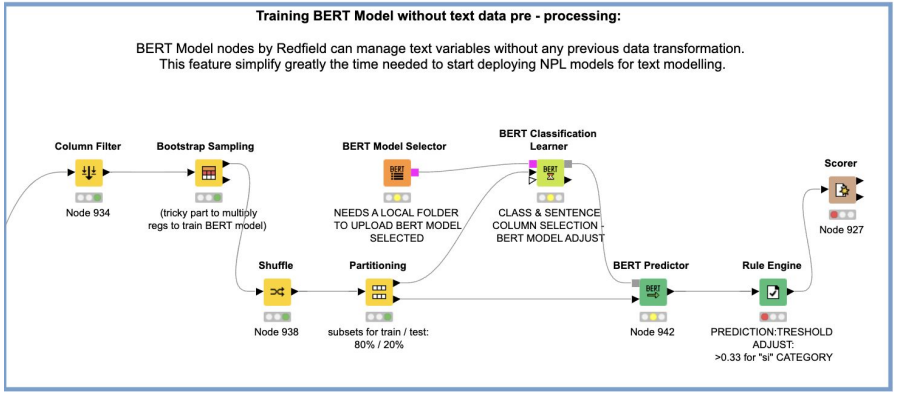


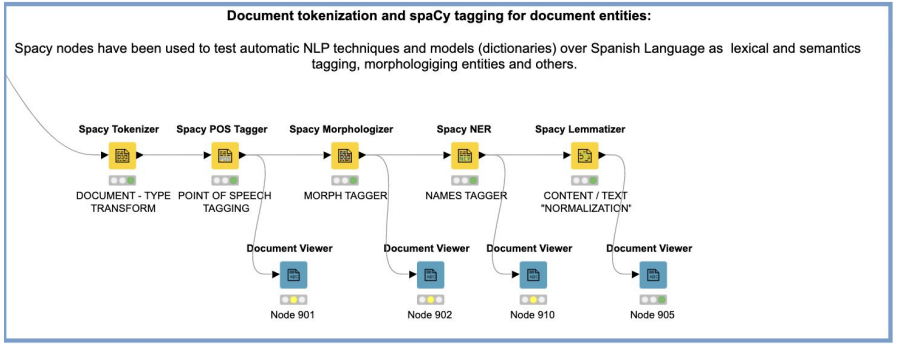
nesta

NLP

Customer opinion analysis with BERT and Spacy

An example project using NLP to analysis customer reviews of online courses using BERT and Spacy





nesta

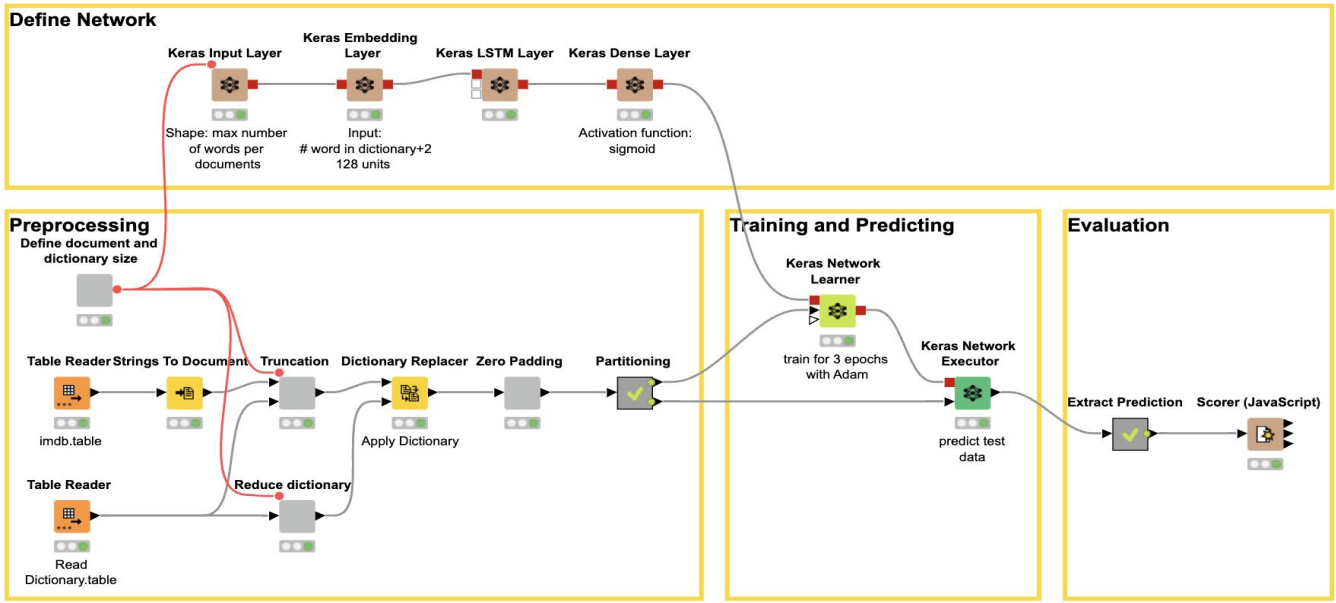
Geospatial analysis

- Harvard Center for Geographic Analysis (CGA) collaboration and geographical analysis related nodes and workflows

nesta

Sentiment analysis using neural networks

This workflow shows how to train a simple neural network for sentiment analysis. It first learns a 128 dimensional embedding and then a LSTM



nesta

Deployment/integrated deployment

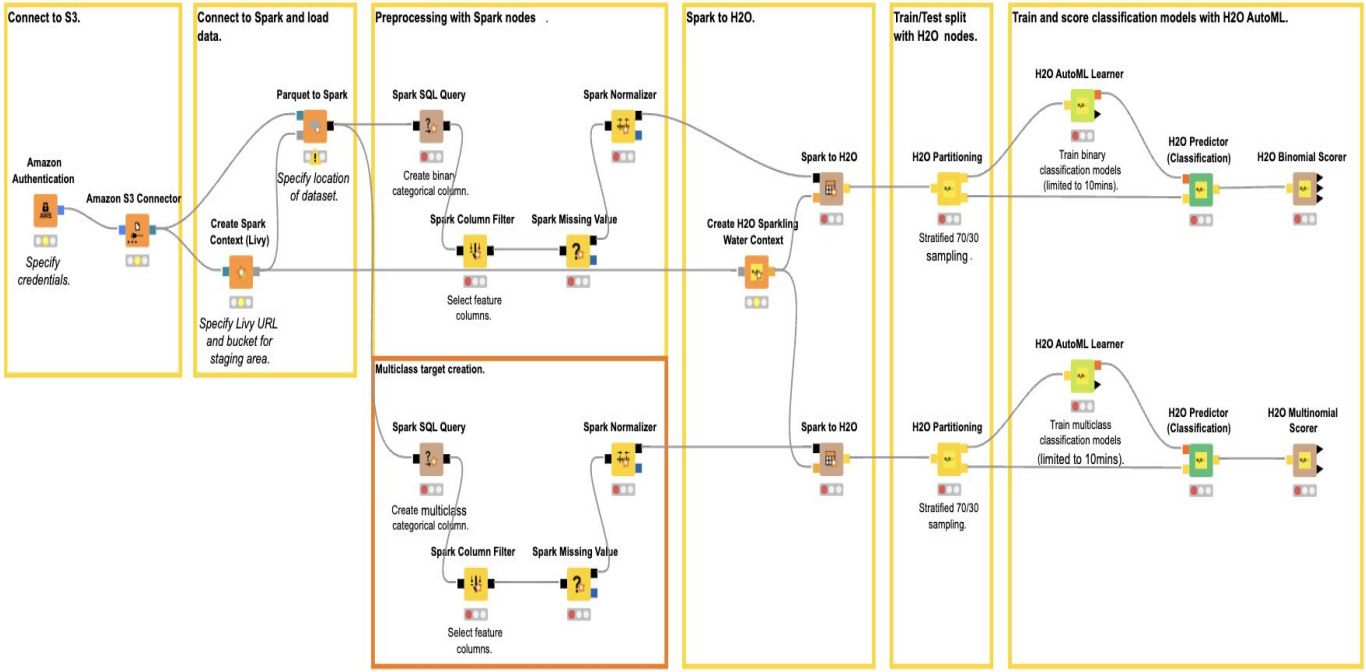
- A series of workflows and blog posts on integrated deployment using KNIME can be found here

- The KNIME Server Rest API

- An example workflow using the Rest API for sentiment analysis

H2O AutoML on Spark

This workflow trains classification models using H2O AutoML on Spark



Python/R & KNIME

- Python guidelines and examples
- Simple example using Python script nodes
- Simple example using a R script to train a model, and reuse it later

nesta

Further reading

Resources

- A few KNIME self-paced courses. A good starting point is the L1-DS course on low-code/no-code for DS and ML
- Beginner examples on KNIME Hub
- A catalogue of learning resources for all learner types
- A blog about low code DS potential in the future
- Events and trainings on low code/no code using KNIME
- KNIME Youtube channel
- An interesting fairness scorer workflow and XAI workflow
- Dataiku functionality

nesta