



Practicing Data Science

A Collection of Case Studies

Rosaria Silipo

Rosaria.Silipo@knime.com



Greetings

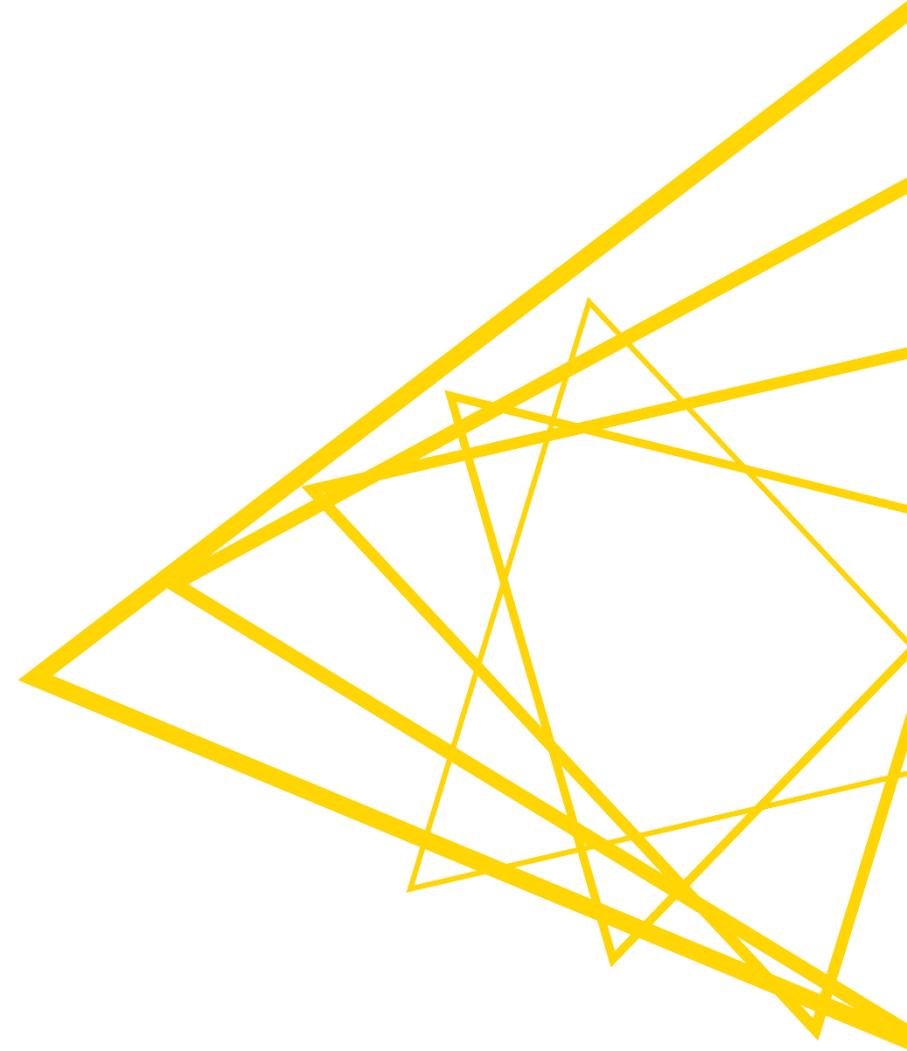
- Rosaria Silipo
- Data Scientist since 1992
- Head of Evangelism at KNIME



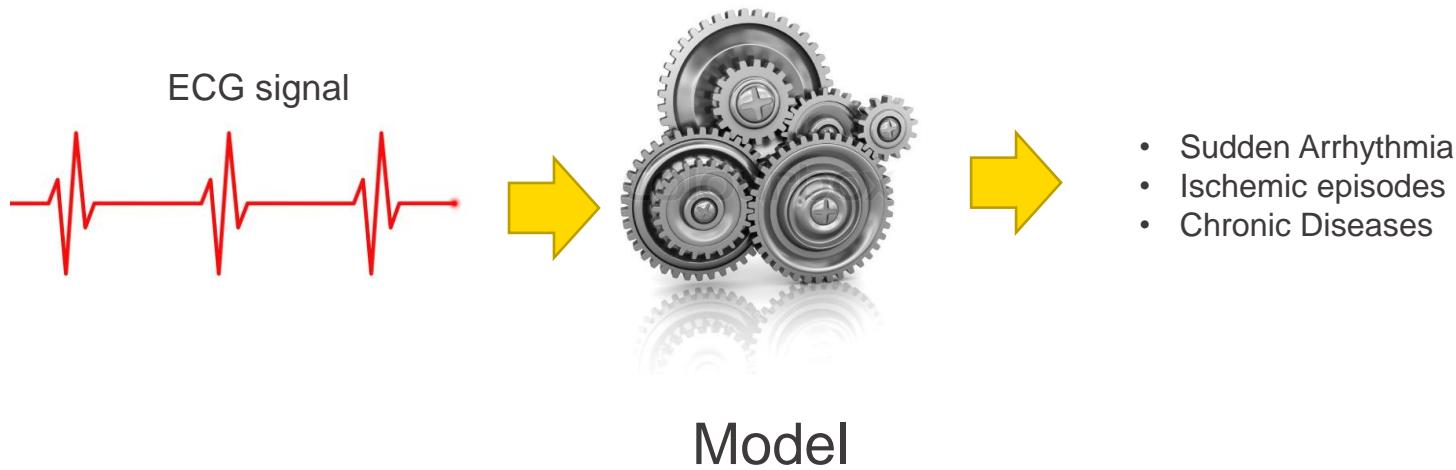
Professional Experience



ECG Classification

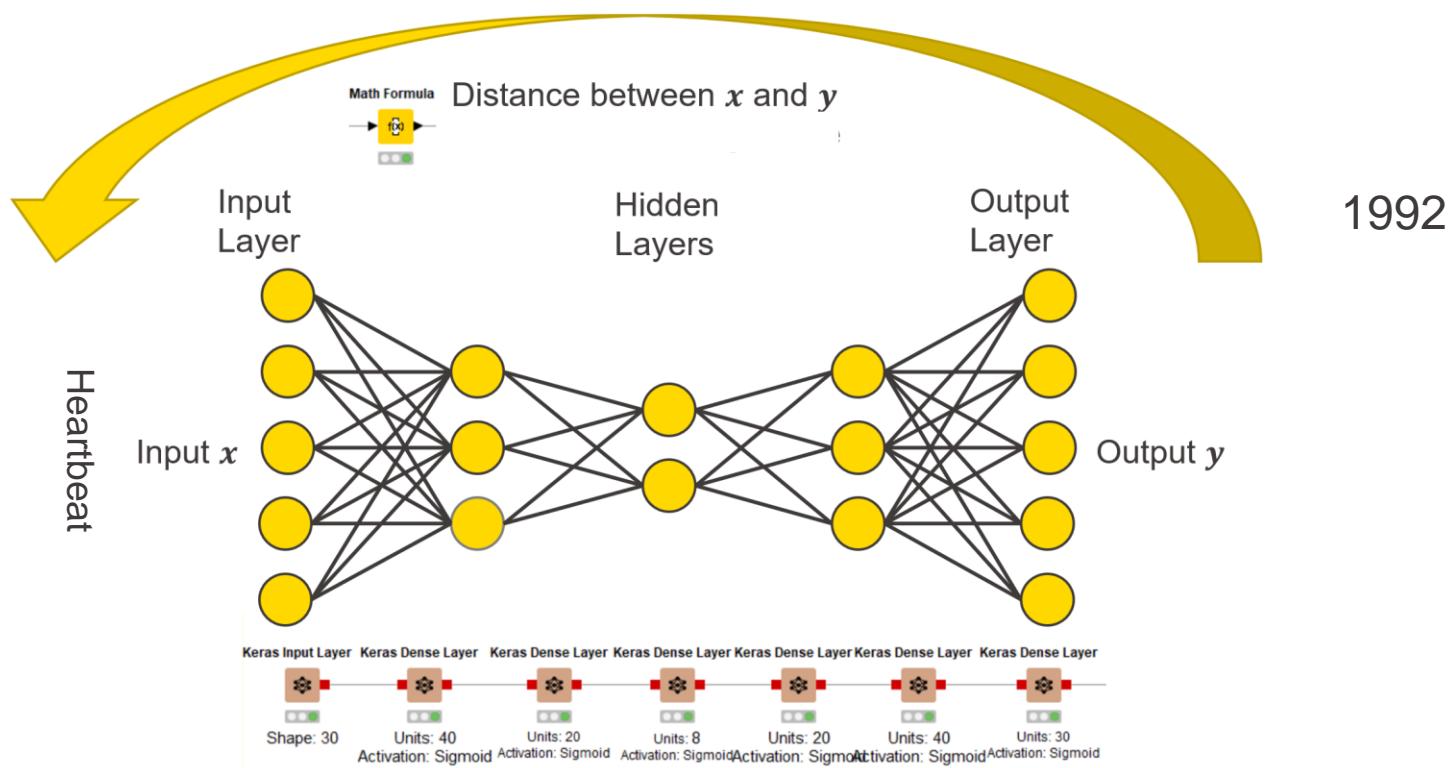


Heart beats in ECG



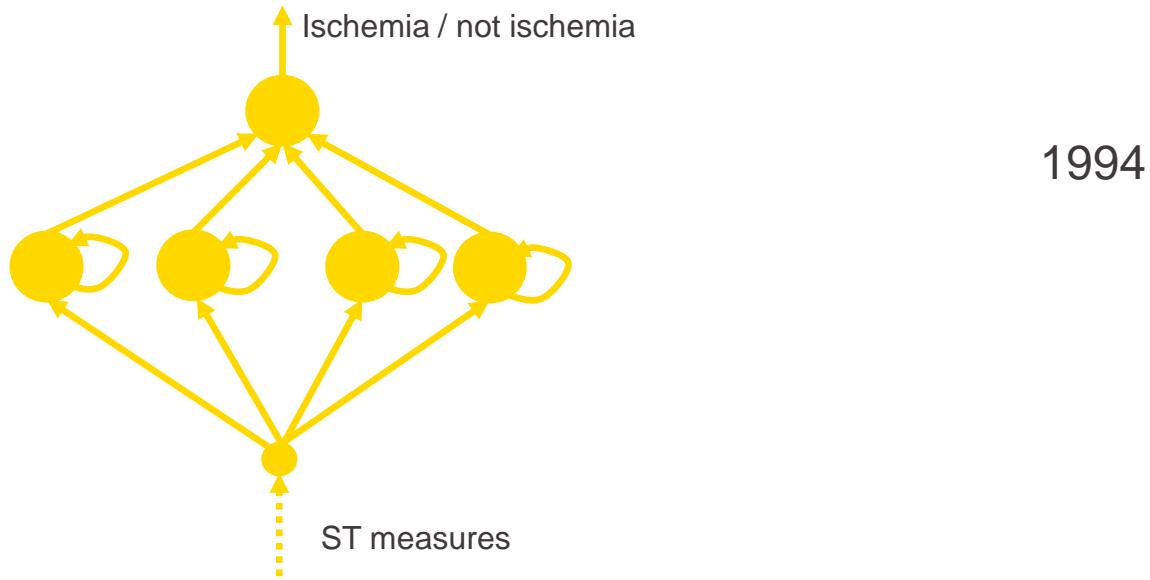
Neural Networks: autoencoder

- Sudden Arrhythmia -> anomaly detection -> neural auto-encoder



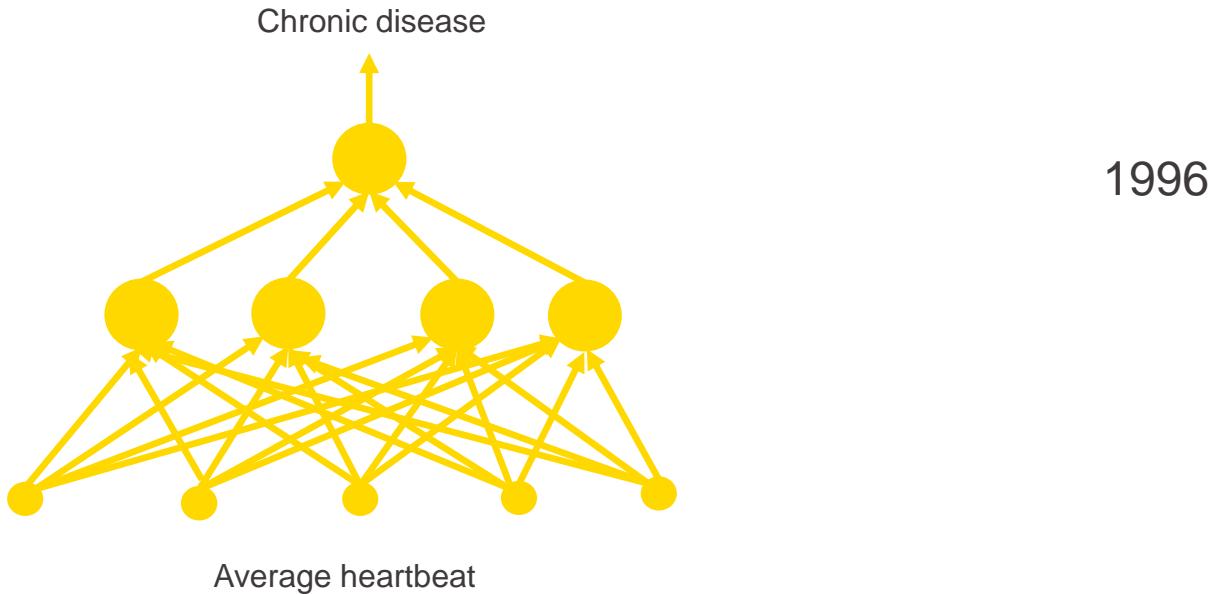
Neural Networks

- Ischemic Episodes -> slow variation over time -> RNNs



Neural Networks

- Chronic Diseases -> QRS shape clax. -> MLPs



Lesson Learned

School vs. work



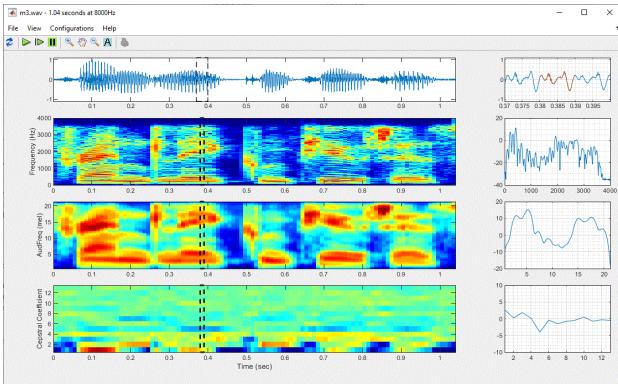
- 17 women out of 300 in first year EE Florence Univ.
- Competition in work environment is tougher

Signal Processing & Speech Recognition



Speech Recognition

speech signal

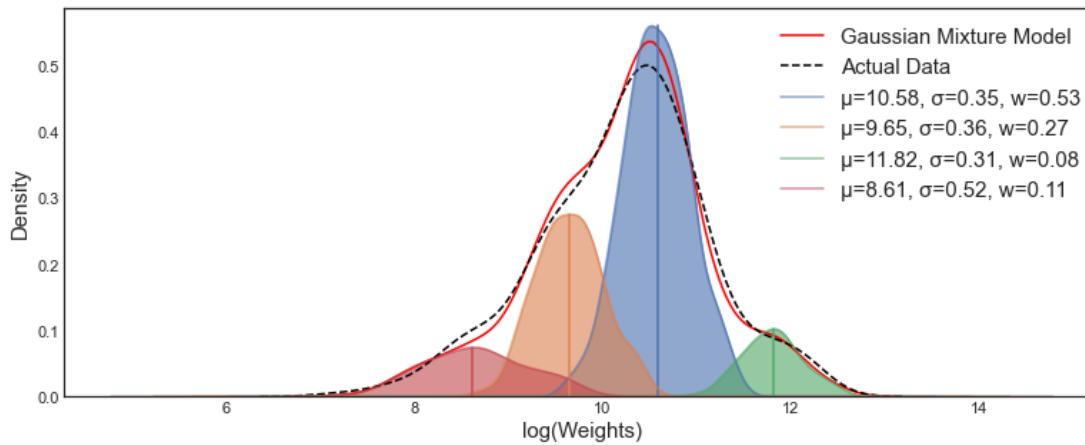


Model

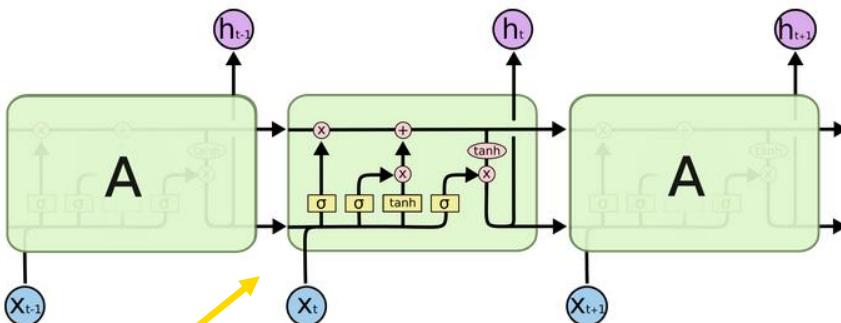
Gaussian Mixture Models

- Hidden Markov Models
- Mixture of Gaussians

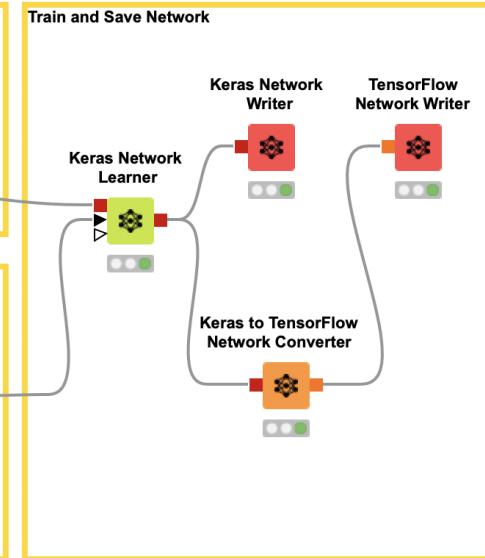
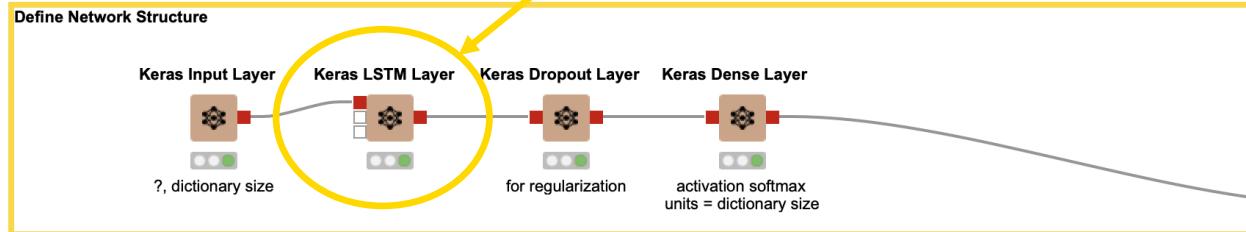
2000



LSTM Networks



2015

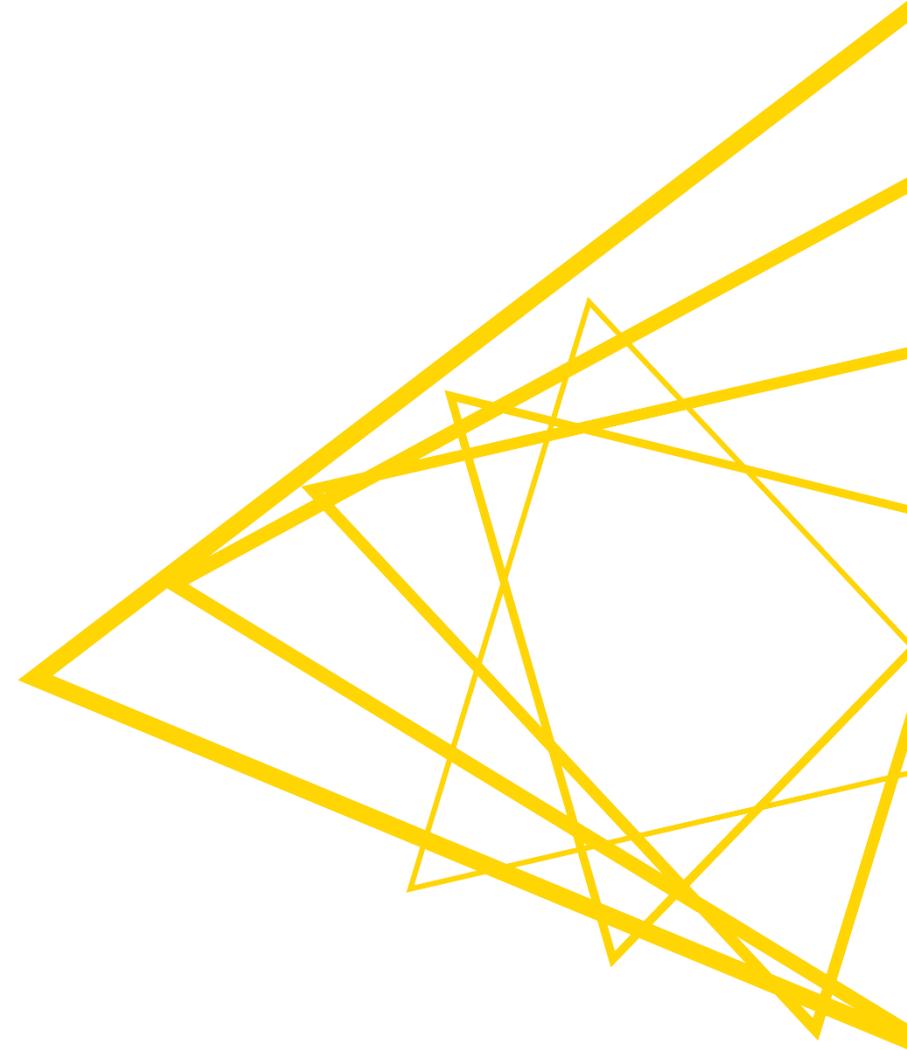


Lesson Learned



- First job is important
- **Great mentors are more important!**

Churn Prediction



Churn Prediction: The Problem

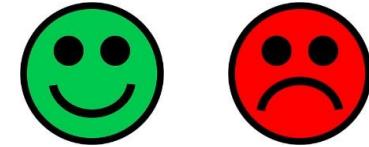


CRM System
Data about your customer

- Demographics
- Behavior
- Revenues

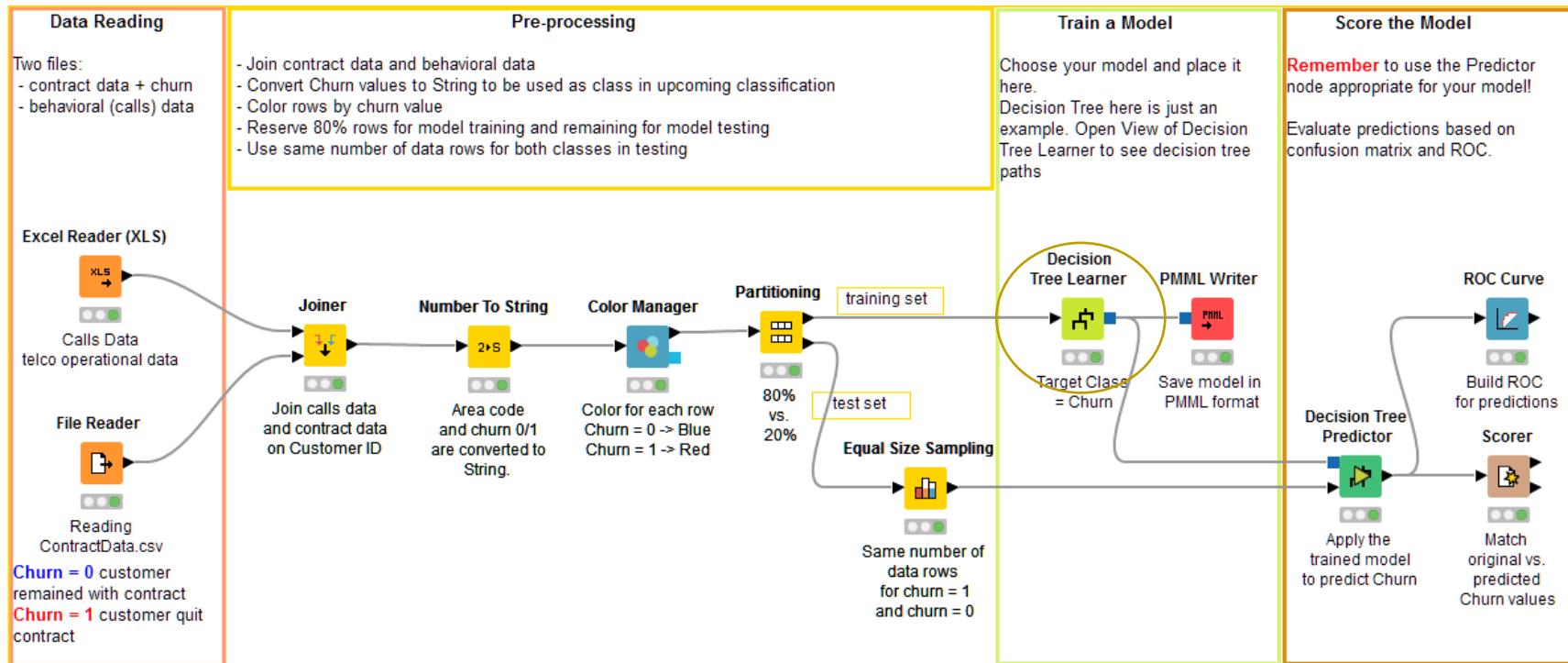


Model



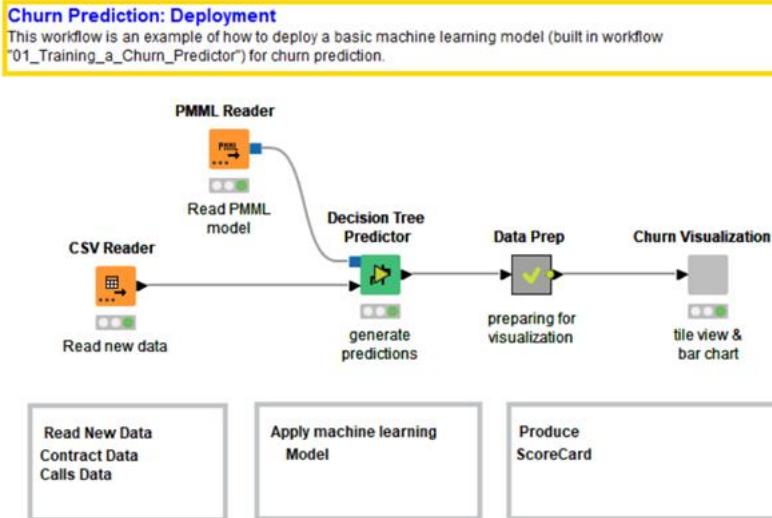
- Churn Prediction
- Upselling Likelihood
- Product Propensity /NBO
- Campaign Management
- Customer Segmentation
- ...

Churn Prediction: The Training Workflow

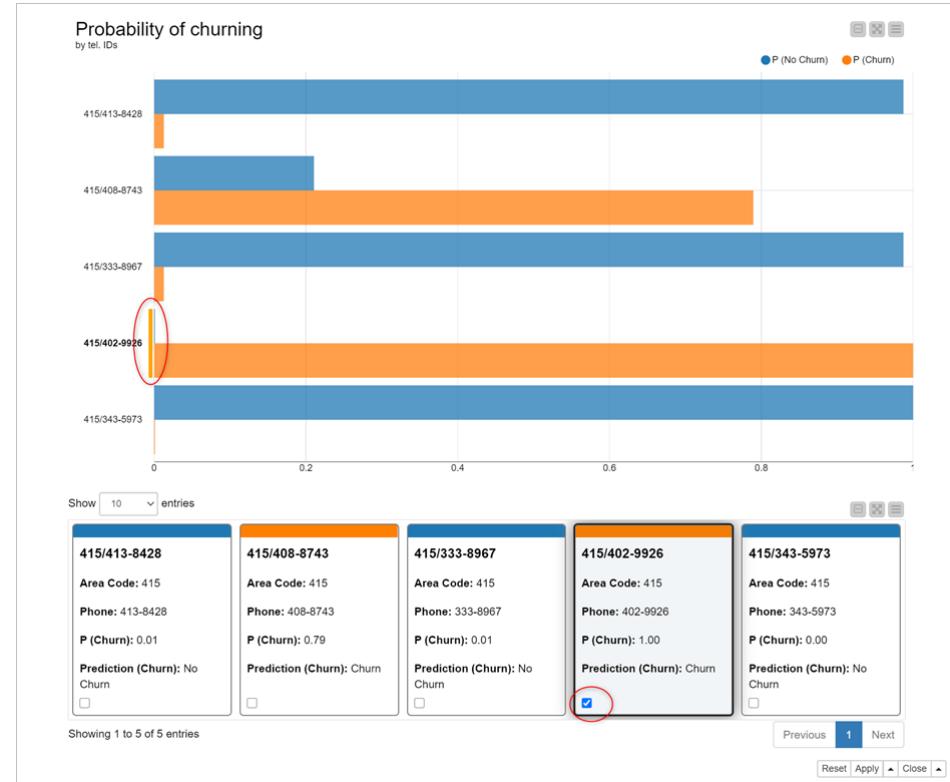


Churn Prediction: The Deployment Workflow

YouTube: "Building a basic Model for Churn Prediction with KNIME" <https://www.youtube.com/watch?v=RHsO10q7e2Y>

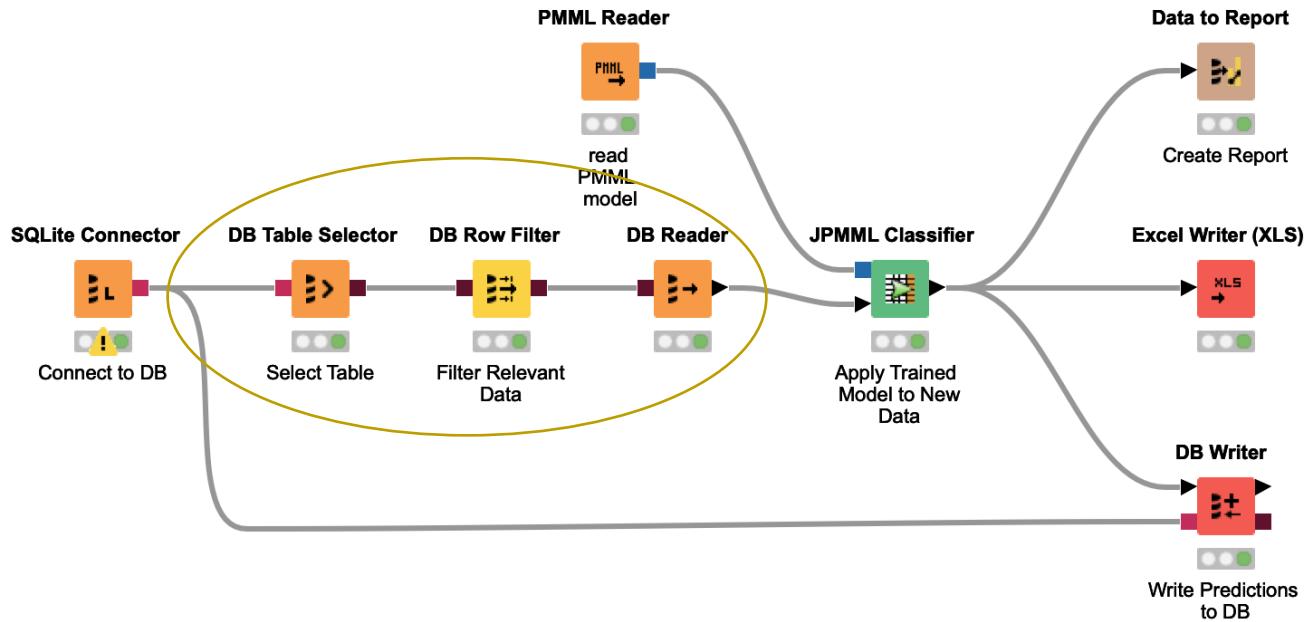


EXAMPLES Server: 50_Applications/18_Churn_Prediction

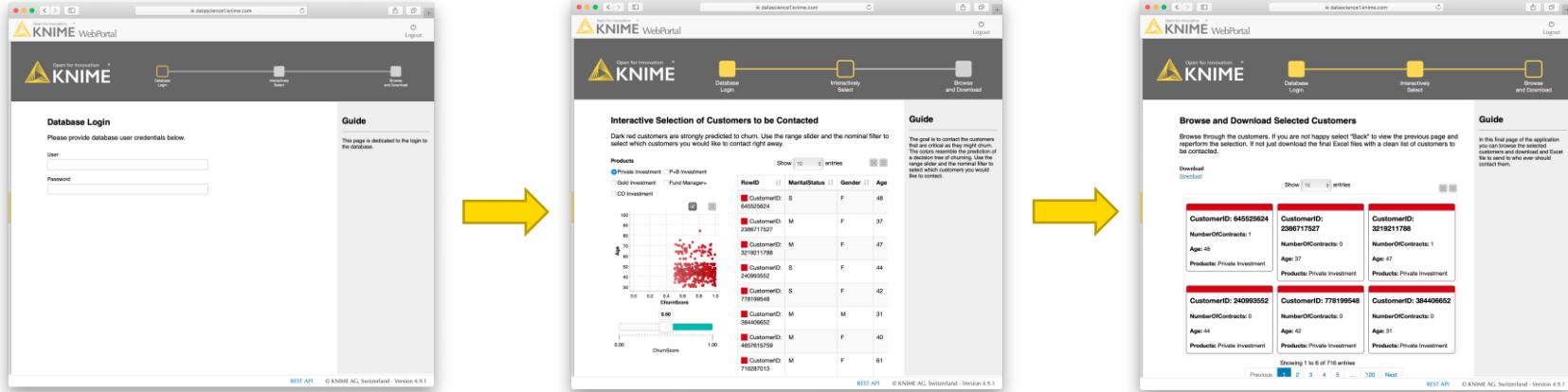


A Second Workflow for Deployment

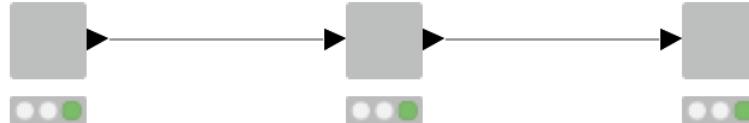
2007



Churn Prediction: Deployment on the WebPortal



Get Customers from Database Select Customers to Contact Browse and Download Customers List



2007

Lesson Learned

- Motherhood and Career



ETL for Banks: ARR Calculation



Annual Recurring Revenues

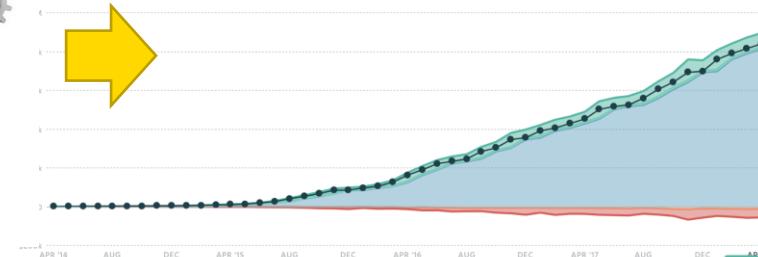
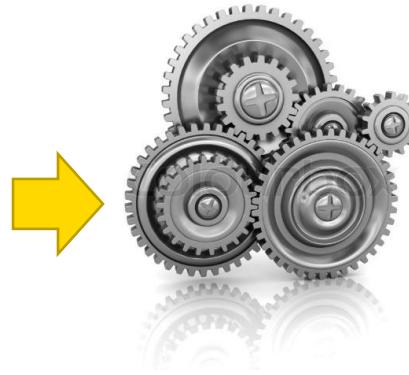
Contracts

File Table - 3:1524 - Excel Reader (Read demo data)

File Edit Hilite Navigation View

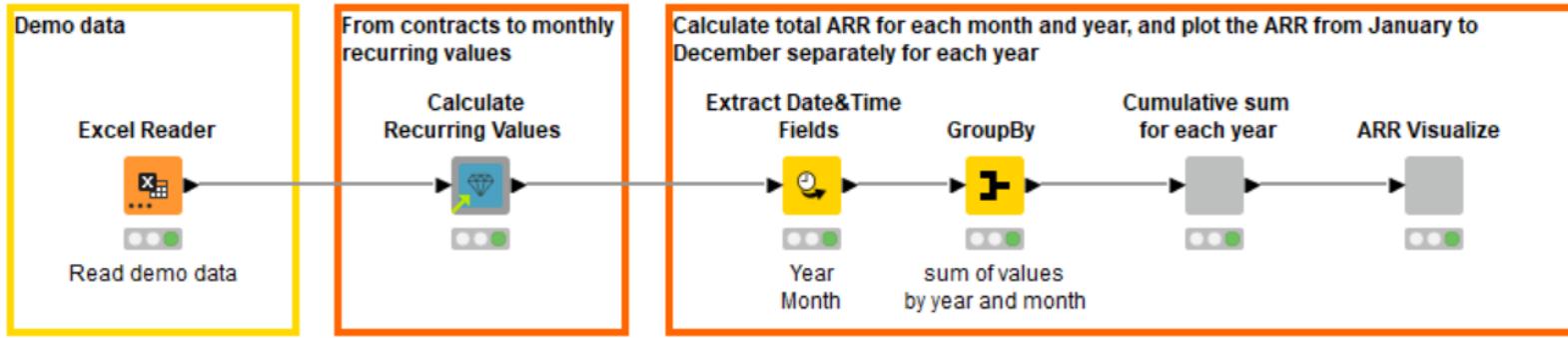
Table "default" - Rows: 45 Spec - Columns: 4 Properties Flow Variables

Row ID	S	CustID	I	Value	Start	End
Row0		Anton		10000	2015-01-01	2015-12-31
Row1		Anton		6700	2016-01-01	2016-08-31
Row2		Anton		2000	2016-09-01	2016-12-31
Row3		Anton		12000	2017-01-01	2017-12-31
Row4		Anton		20000	2018-01-01	2018-12-31
Row5		Anton		20000	2019-01-01	2019-12-31
Row6		Horst		5000	2015-07-01	2015-12-31
Row7		Horst		5000	2016-01-01	2016-06-30
Row8		Horst		2000	2016-07-01	2016-12-31
Row9		Horst		4000	2017-01-01	2017-06-30
Row10		Horst		6000	2017-07-01	2017-12-31
Row11		Horst		12000	2018-01-01	2018-12-31
Row12		Horst		12000	2019-01-01	2019-12-31
Row13		Otto		4000	2015-09-01	2015-12-31
Row14		Otto		8000	2016-01-01	2016-08-31
Row15		Thomas		7000	2016-01-01	2016-06-30
Row16		Thomas		10000	2016-07-01	2016-12-31
Row17		Thomas		17000	2017-01-01	2017-12-31
Row18		Thomas		10000	2018-01-01	2018-12-31
Row19		Thomas		10000	2019-01-01	2019-12-31
Row20		Thomas		2000	2016-04-01	2016-12-31



Model

Annual Recurring Revenues



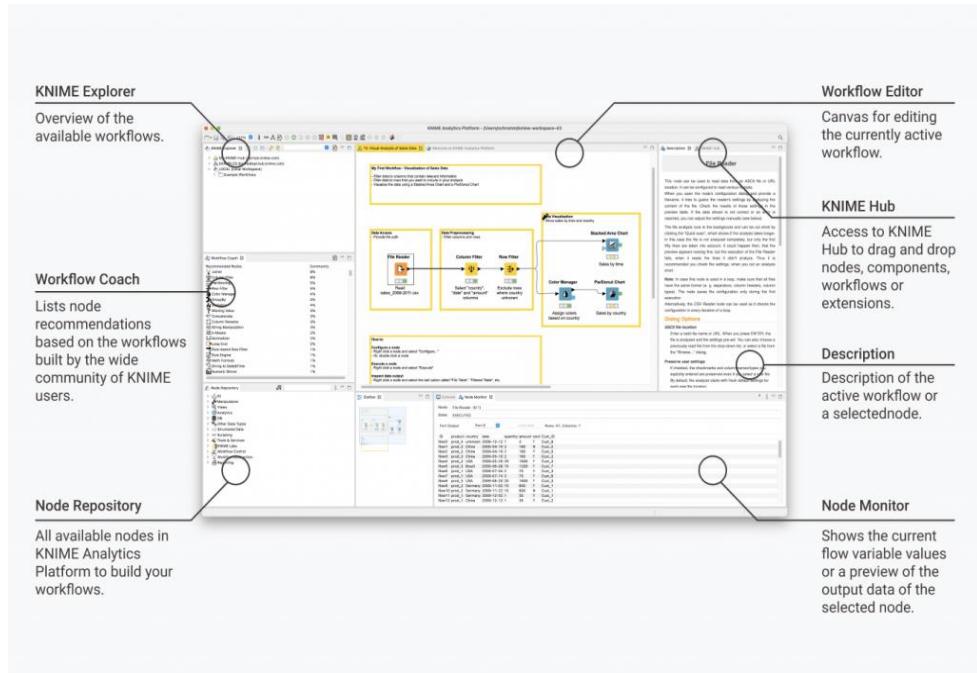
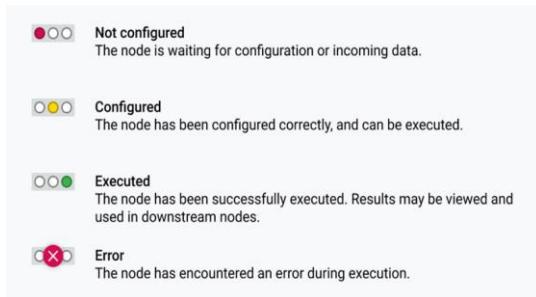
2009

ARR final page



Lesson Learned

- This is when I first got to know KNIME
- Open Source
- Low Code
- Large Coverage
- Active Community
- Examples on KNIME Hub
- Support on KNIME Forum

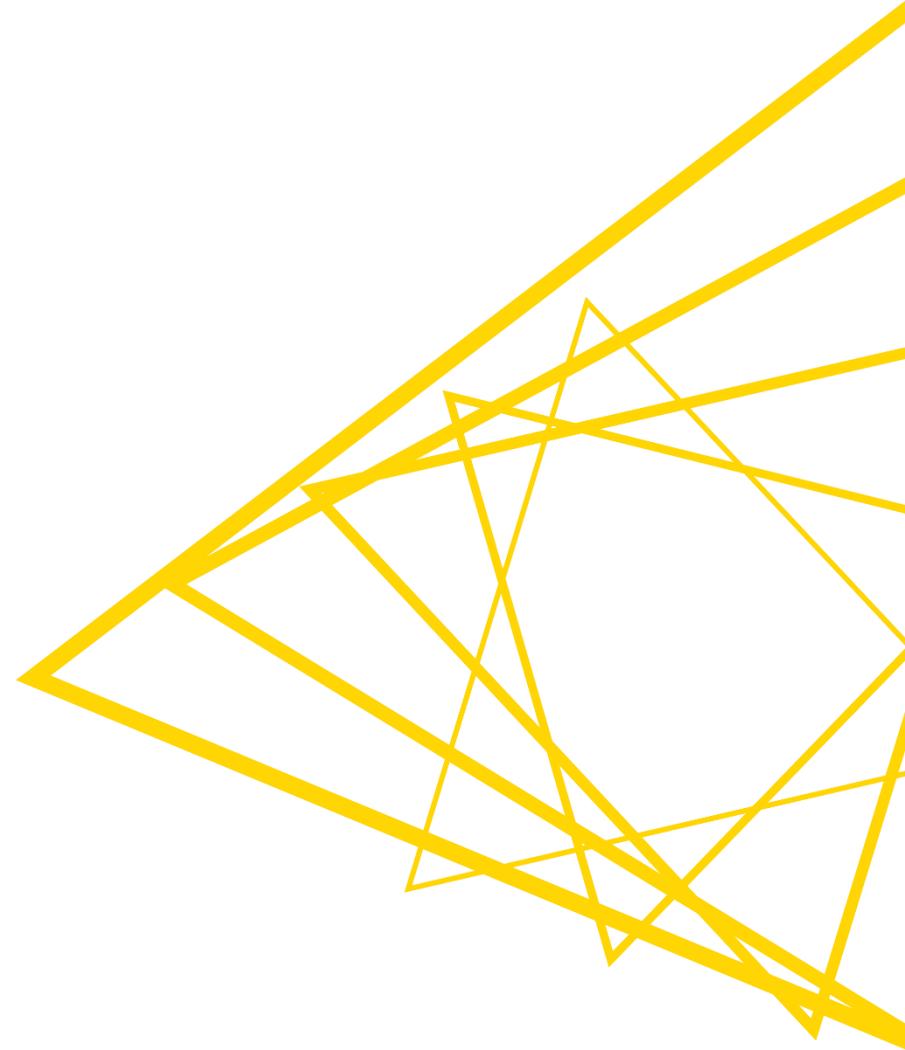


KNIME Hub (<https://hub.knime.com>)

The screenshot shows a web browser window displaying the KNIME Hub search results. The URL in the address bar is <https://hub.knime.com/search?q=product%20naming%20LSTM&type=Workflow>. The search term "product naming LSTM" is entered in the search bar. The results page shows 881 results. The "Workflows" tab is selected. Two workflows are listed:

- Generate Text Using a Many-To-One LSTM Network (Deployment)**
The workflow generates text in fairy tale style. It reads the previously trained TensorFlow network and predicts a sequences of index-encoded characters within a loop, and translates the sequence of i...
knime > Examples > 04_Analytics > 14_Deep_Learning > 02_Keras > 11_Generate_Fairy_Tales > 02_Deployment
Tags: deep learning, keras, text generation, RNN, LSTM, text analysis, sequence analysis, neural network, text processing, NLP, Natural Language Processing
- Generate Text Using a Many-To-One LSTM Network (Training)**
The workflow builds, trains, and saves an RNN with an LSTM layer to generate new fictive fairy tales. The brown nodes define the network structure. The "Pre-Processing" metanode reads fairy tales and ...
knime > Examples > 04_Analytics > 14_Deep_Learning > 02_Keras > 11_Generate_Fairy_Tales > 01_Training
Tags: deep learning, keras, text generation, RNN, LSTM, text analysis, sequence analysis, neural network, text processing, NLP, Natural Language Processing

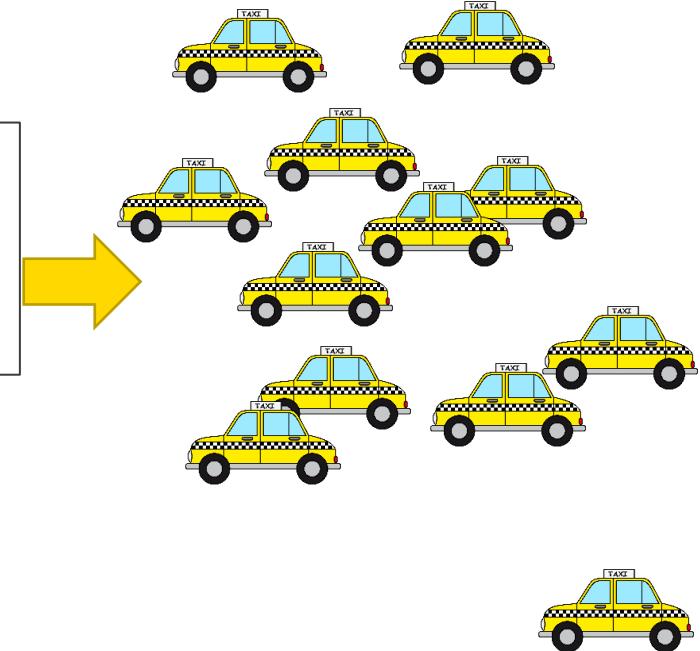
Demand Prediction (Taxi)



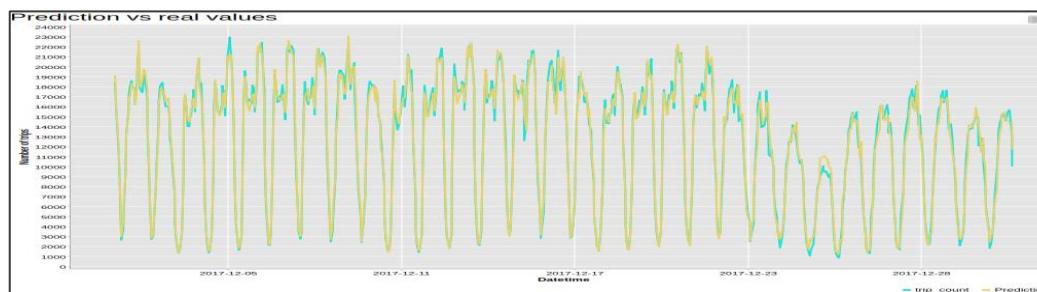
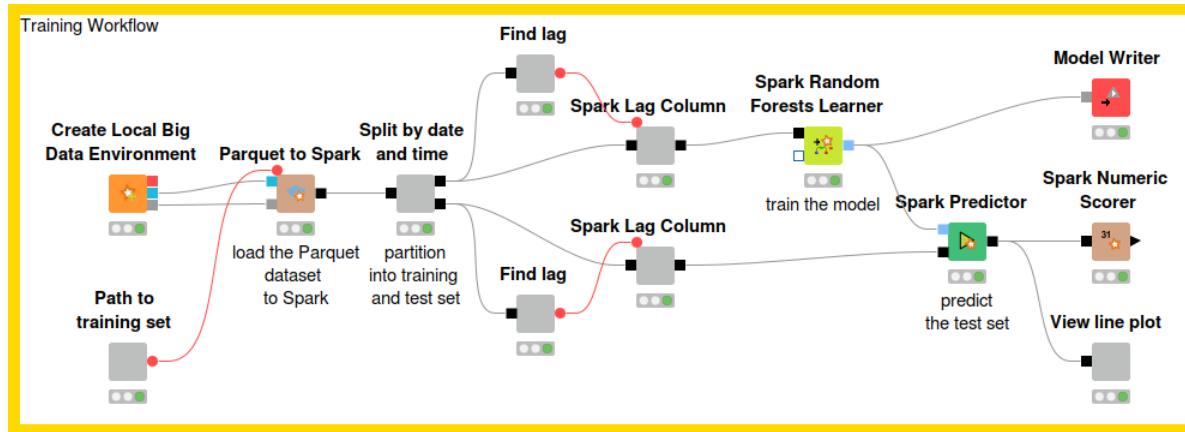
Demand Prediction: The Problem



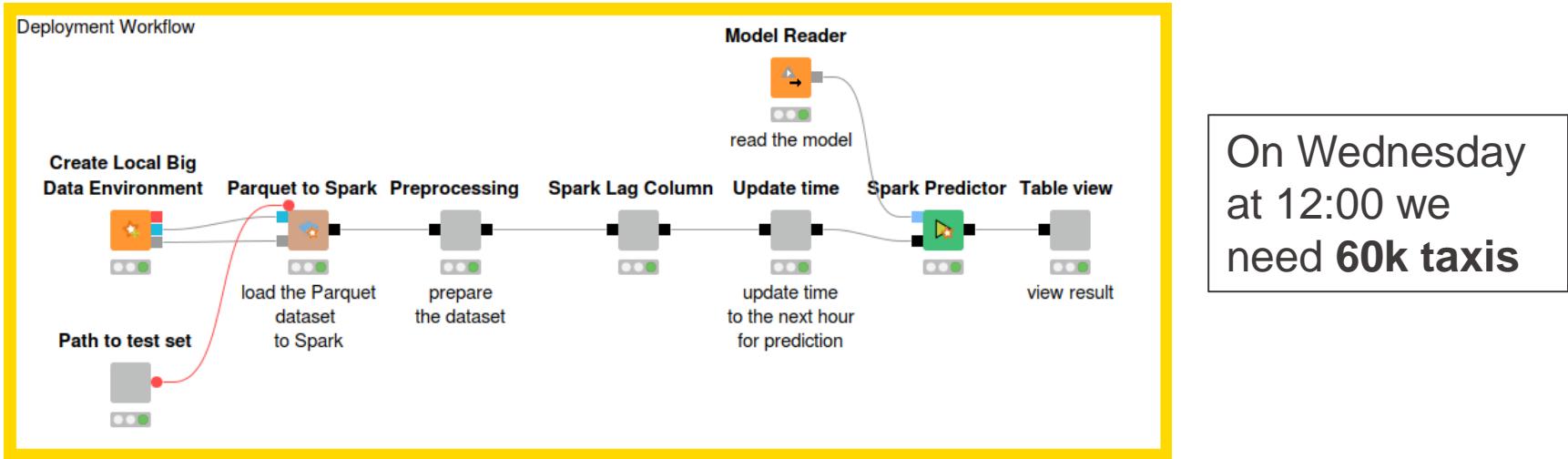
How many taxi do
I need in NYC on
Wednesday at
12:00?



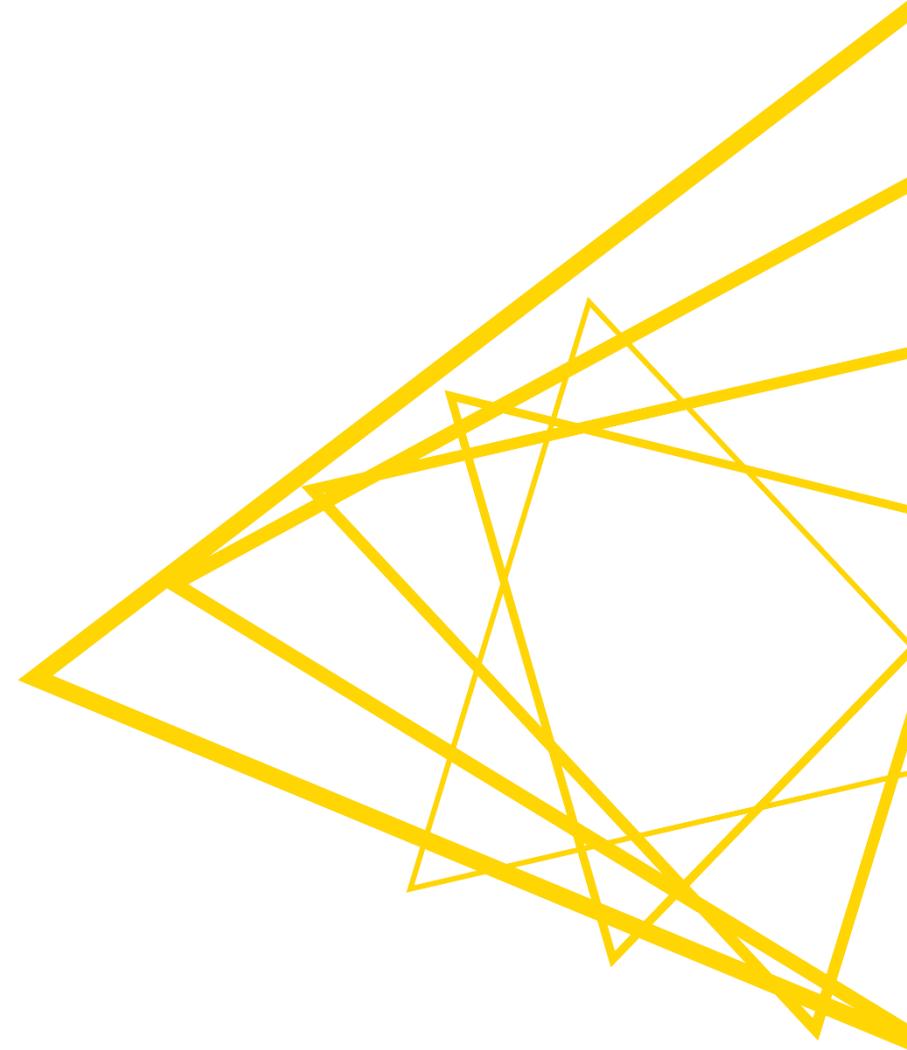
Demand Prediction: The Training Workflow



Demand Prediction: Deployment



Recommendation Engines



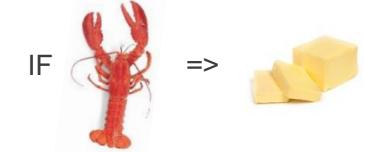
Recommendation Engines or Market Basket Analysis



Model

$$\text{Support} = \frac{\text{frq}(X, Y)}{N}$$
$$\text{Rule: } X \Rightarrow Y \quad \text{Confidence} = \frac{\text{frq}(X, Y)}{\text{frq}(X)}$$
$$\text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}$$

Recommendation



Inspired by your purchases



theory11 Artisan Playing Cards (White)
★★★★★ 152
\$10.75



theory11 Artisan Playing Cards (Black)
★★★★★ 71
\$9.60



theory11 High Victorian Playing Cards
★★★★★ 15
\$10.70



theory11 Citizen Playing Cards
★★★★★ 72
\$9.93 prime
\$30.46



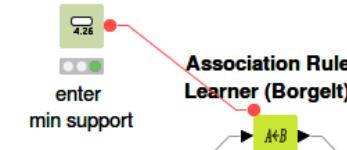
The Poetry and Short Stories of Dorothy Parker
• Dorothy Parker
★★★★★ 18
Hardcover
\$30.46

Market Basket Analysis: with Association Rules

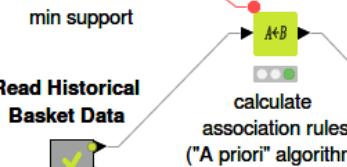
Market Basket Analysis: Build Association Rules

1. Read Transaction/Basket data and Product data
2. Using "A priori" algorithm, build association rule set
 - min. set size = 1
 - min rule confidence = 10%
 - min support is controlled by Double Input Quickform node in %
3. Translate Antecedent collections into product name concatenations
4. Translate Consequent Item ID into Consequent Product Name
5. Calculate price stats and rule revenue
6. Write association rule set to file

Double Input

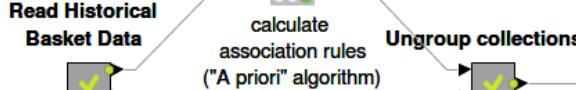


Association Rule Learner (Borgelt)



Read Historical Basket Data

- Read data
1. Transactions
2. Products



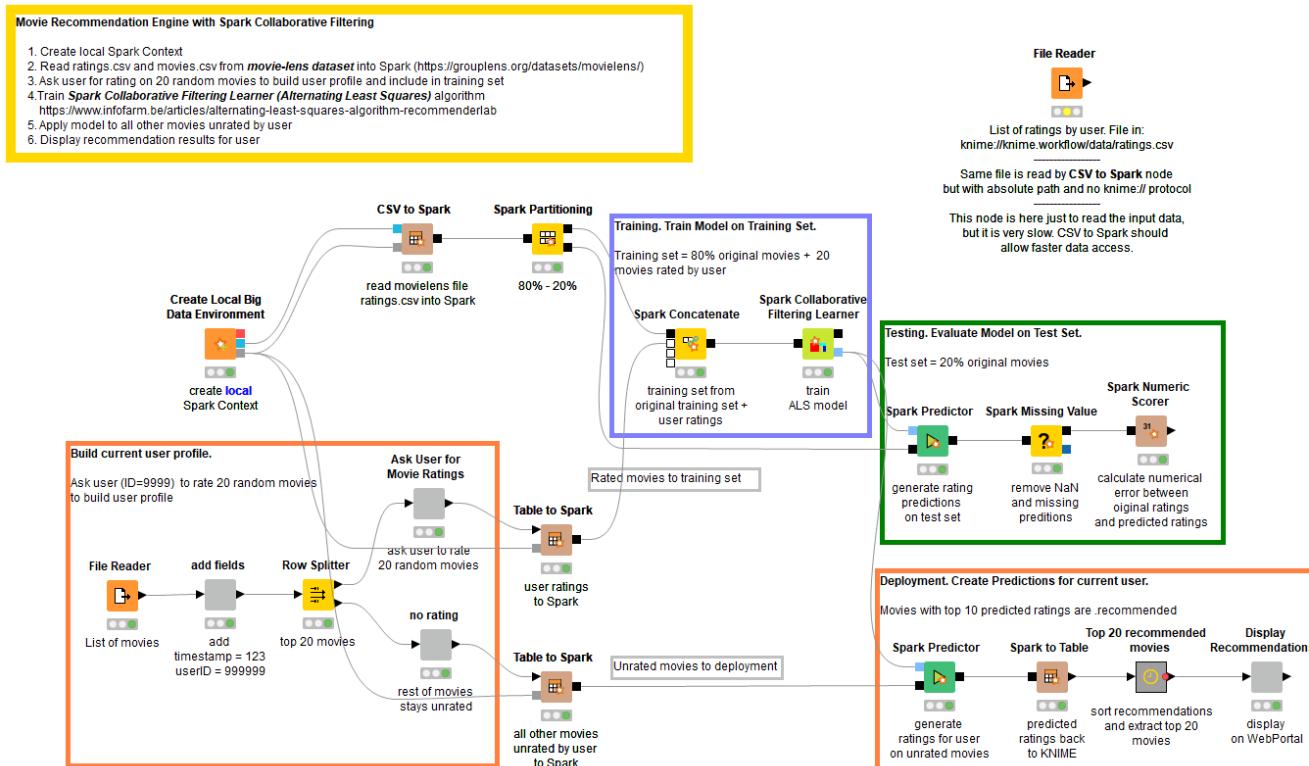
Ungroup collections

- translate antecedent collection to product name concatenations
- associate consequent product ID to product name
- calculate price stats and rule revenue

Table Writer

- write association rules to a file

Recommendation Engine: with Collaborative Filtering



Recommendation Engine/MBA: Deployment

Basket Analysis Report

Welcome to our Supermarket Chain!

The total price for your current shopping cart is **79.17\$!**

Purchase Advices

1. Try our **lobster** !

Today's price for lobster is just 23.72\$!

... and if you like our **shrimps**, we are sure you will also enjoy the **lobster** !

2. Try our **lobster** !

Today's price for lobster is just 23.72\$!

... and if you like our **cookies**, we are sure you will also enjoy the **lobster** !



Created with KNIME Report Designer. Provided by KNIME.com AG, Zurich, Switzerland

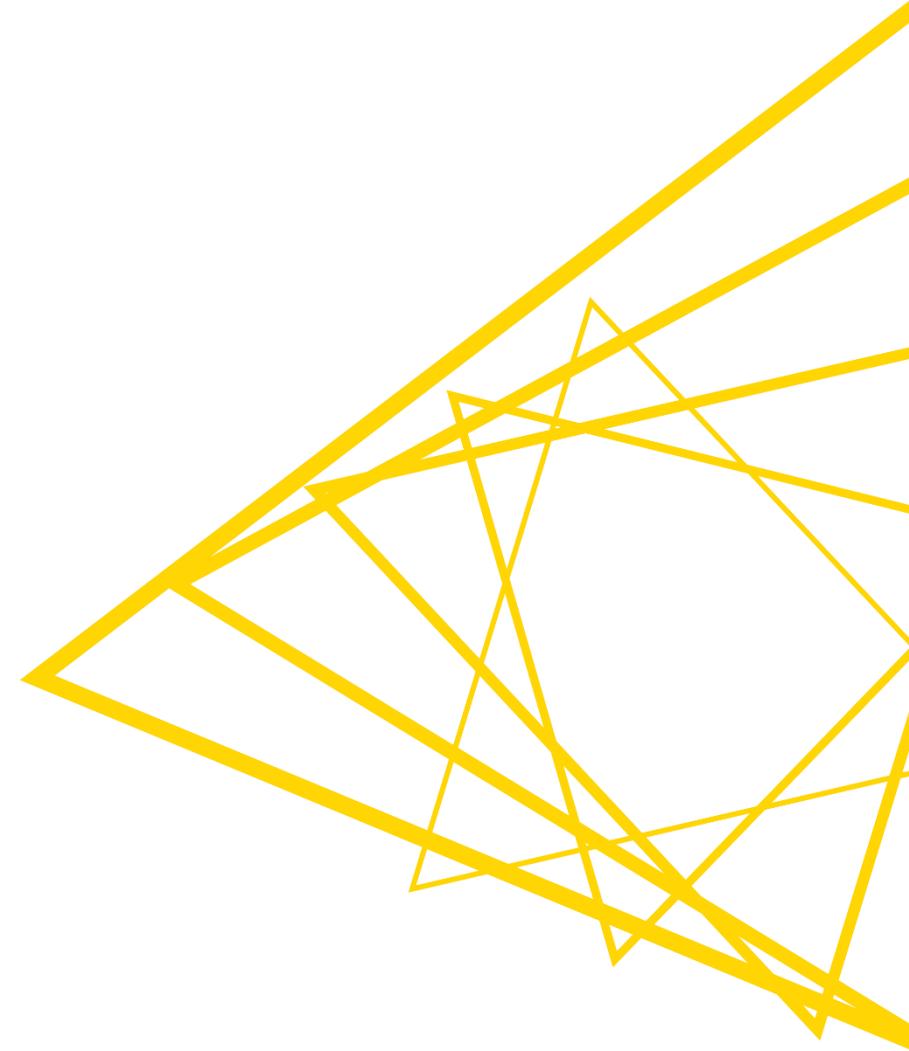


Lesson Learned



- Find a job where you can learn ...

Fraud Detection

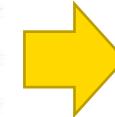
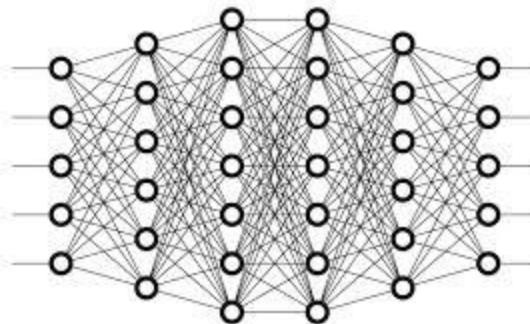


Fraud Detection



Transactions

- Trx 1
- Trx 2
- Trx 3
- Trx 4
- Trx 5
- Trx 6
- ...



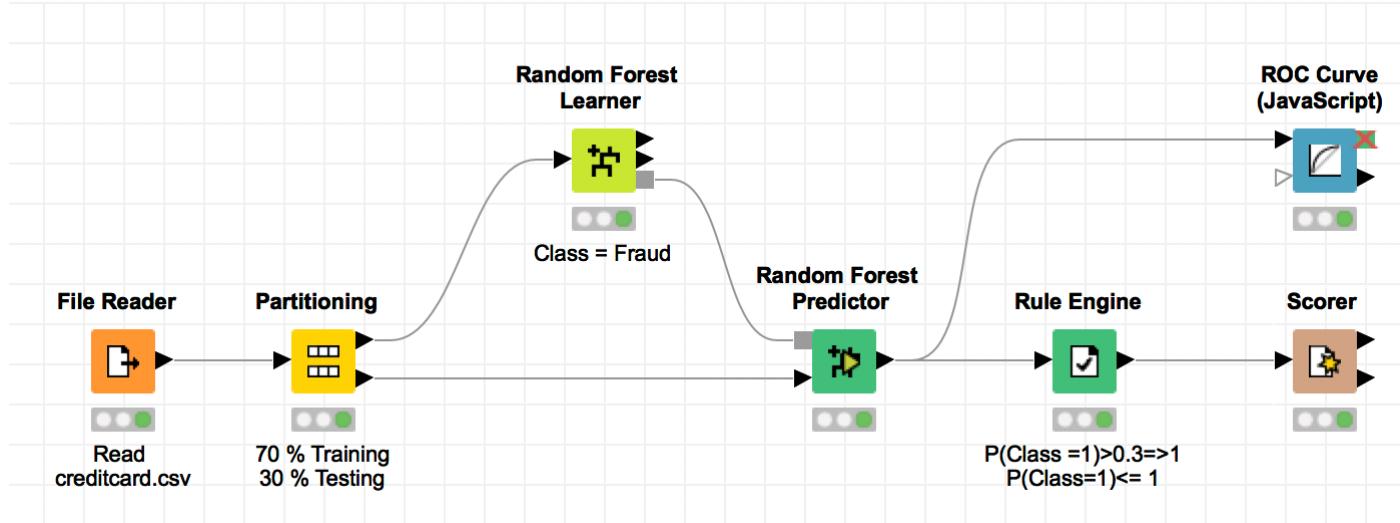
- Good
- Good
- Good

• **Fraud**

- Good
- Good
- ...

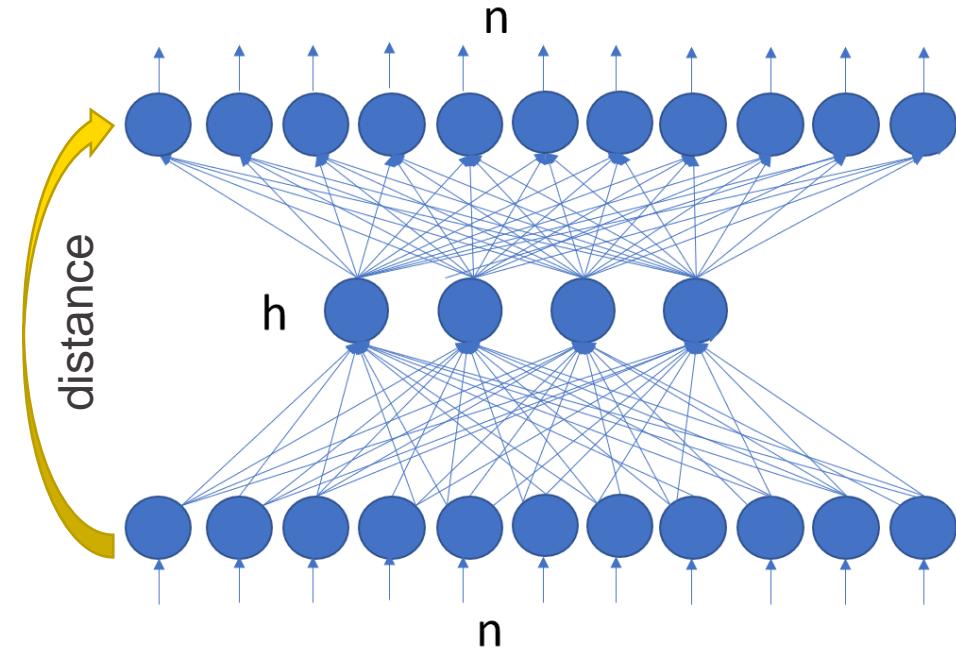
Model

Fraud Detection with Examples

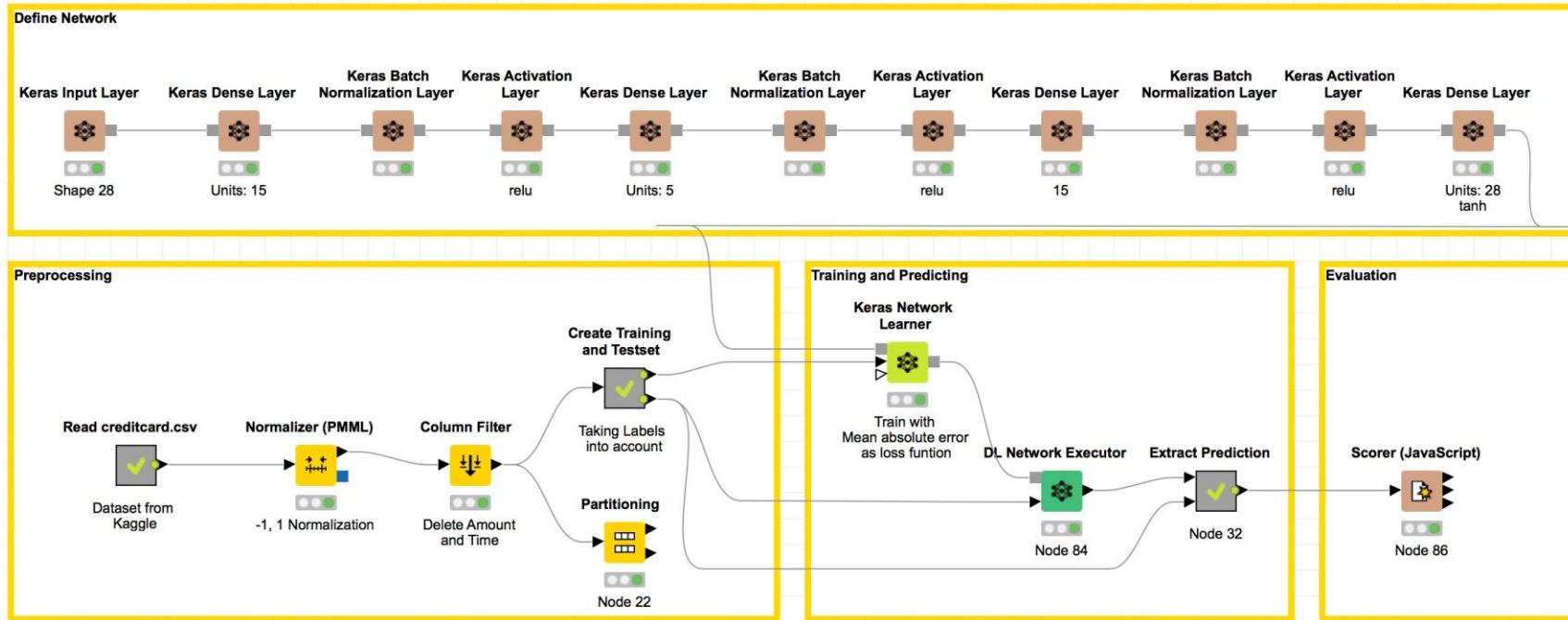


Fraud Detection: without Examples – Auto-encoder

- Trained with Back-Propagation on just “normal” transactions
- If $\text{distance} > \text{threshold}$
=> possible fraud

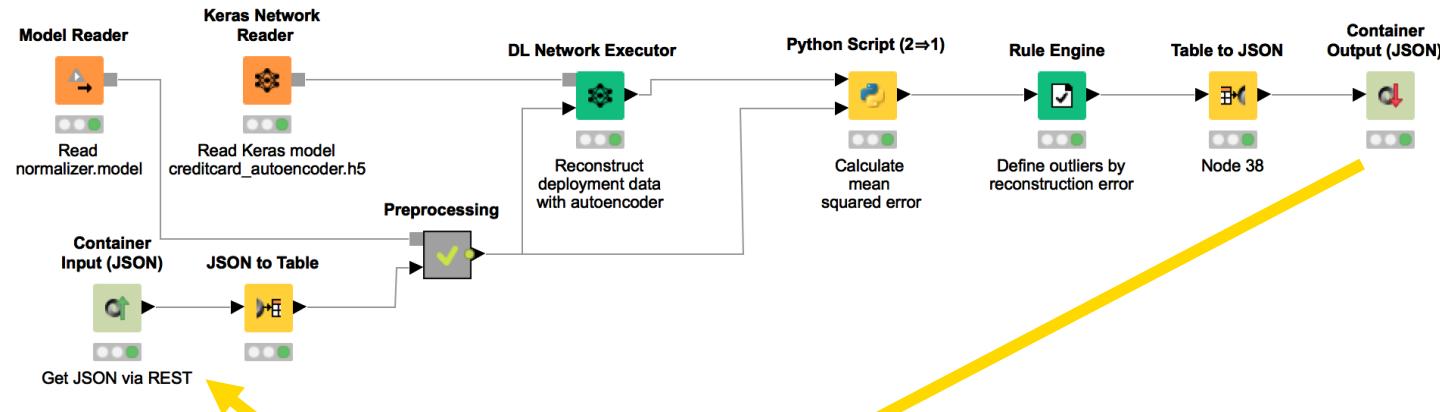


Fraud Detection without Examples

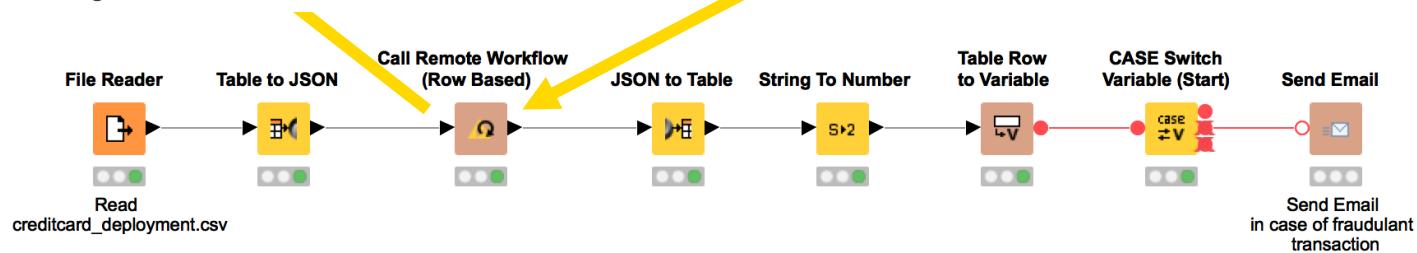


Deployment via REST on KNIME Server

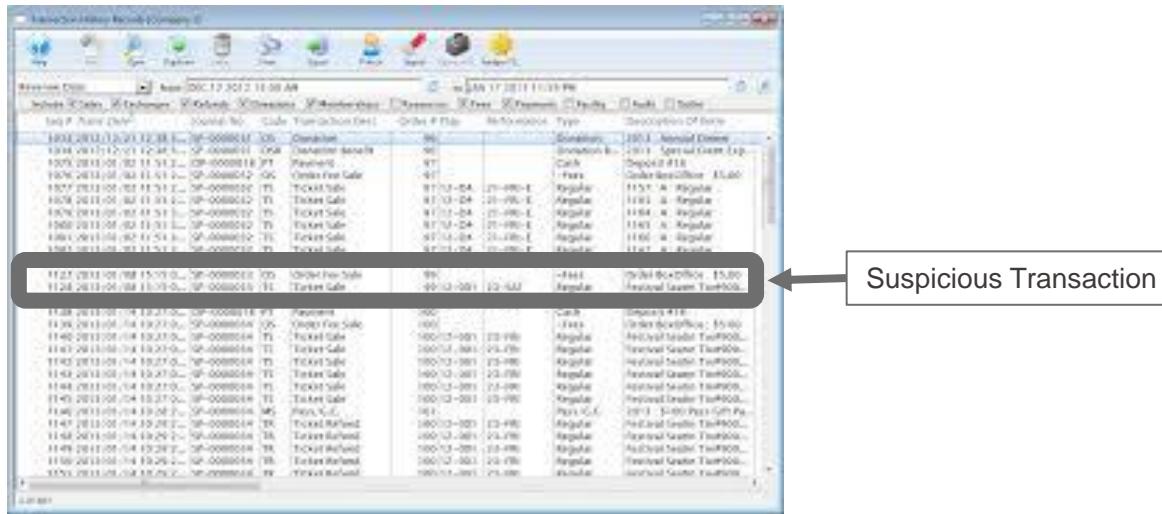
Workflow deployed as (REST) web service on KNIME Server



Workflow calling another workflow on KNIME Server



Fraud Detection deployed



Anomaly Detection in Predictive Maintenance (IoT)



Anomaly Detection: The Problem

Dynamic Unsupervised Anomaly Detection

Some measures change over time till their values are not normal anymore. For example, while a motor is slowly deteriorating, one of the measurements might change till it gets out of control and the motor breaks. We want to stop the motor before it completely breaks producing even more damages

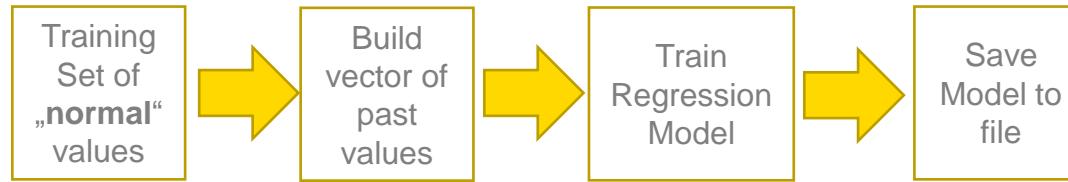
The Data: Time Series

- 28 time series from 28 sensors on 8 different parts of a mechanical engine.
- Time Series are FFT-derived Spectral Amplitudes
- There is only one motor breakdown episodes on July 21, 2008
- The breakdown is visible only from some sensors and only in some frequency bands
- The rotor was substituted with a new one after the breakdown

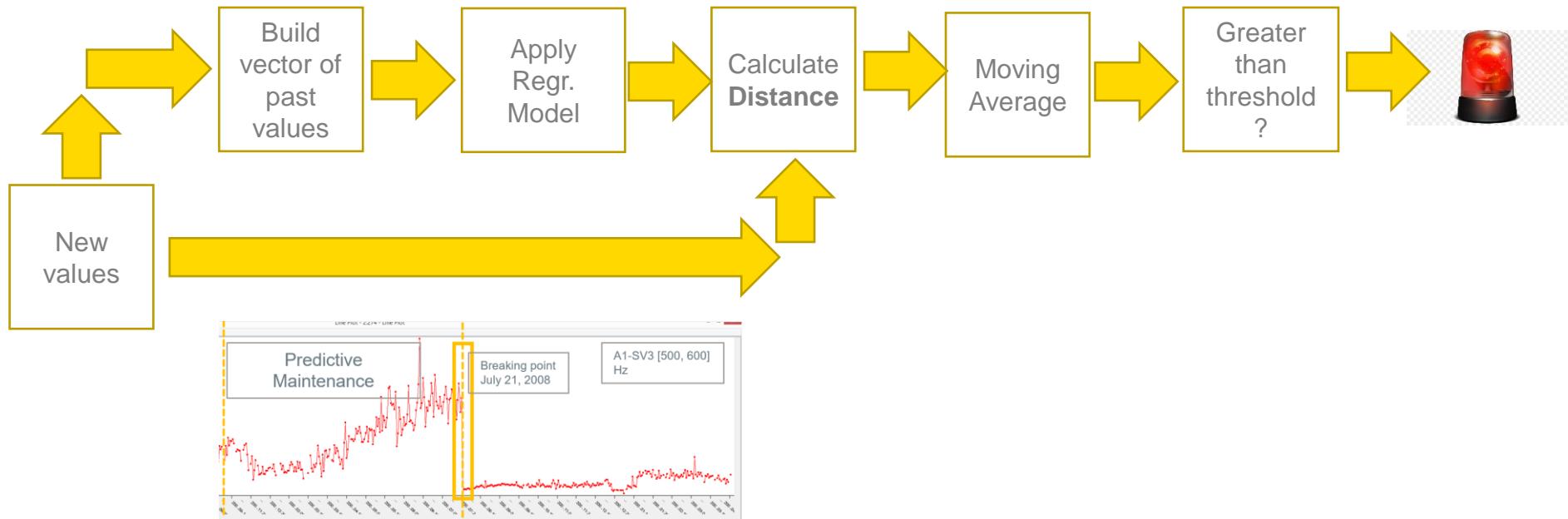
The Data: Time Series



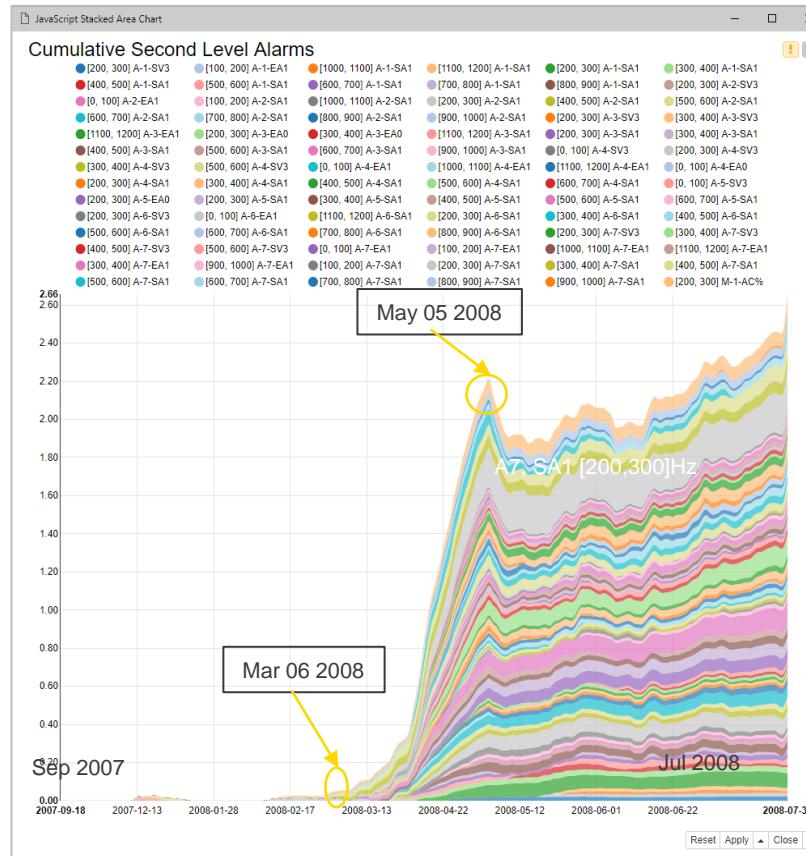
Anomaly Detection: Training



Anomaly Detection: Deployment

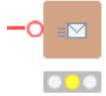


Anomaly Detection: Deployment of Second Level Alarm



Anomaly Detection: Deployment Triggers

Send Email



Dialog - 5:270:269:284:271 - Send Email

To: rosaria.silipo@knime.com

CC:

BCC:

Subject: 'nd Level Alarm detected!

prediction
Mean(alarm(error)[600, 700]+Amp A-3-SA1)
RowID
currentIteration
currentColumnName
knime.workspace

Second Level Alarm has been detected today
in Sensor and Frequency Band \${currentColumnName}\$\$

Your Motor Administrator

Priority: Normal Text HTML

OK Apply Cancel ?



via REST

Lesson Learned

- Do women have less credibility as data scientist?
- When they see how good I am ...

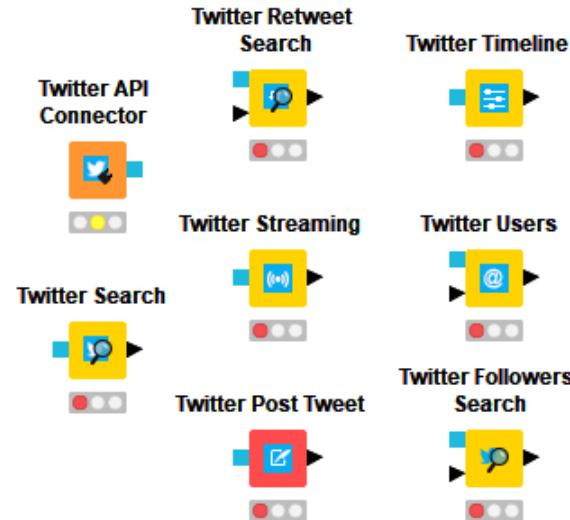


Mining Social Media Activity from Twitter

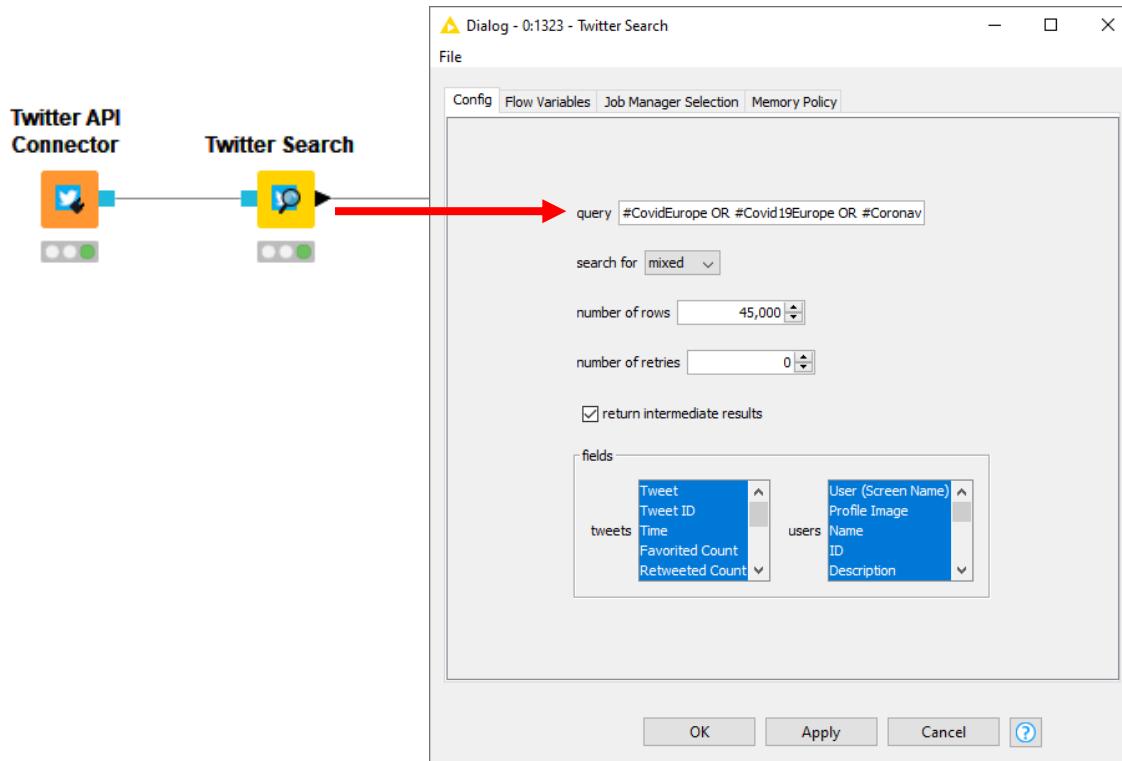


The Twitter API Integration

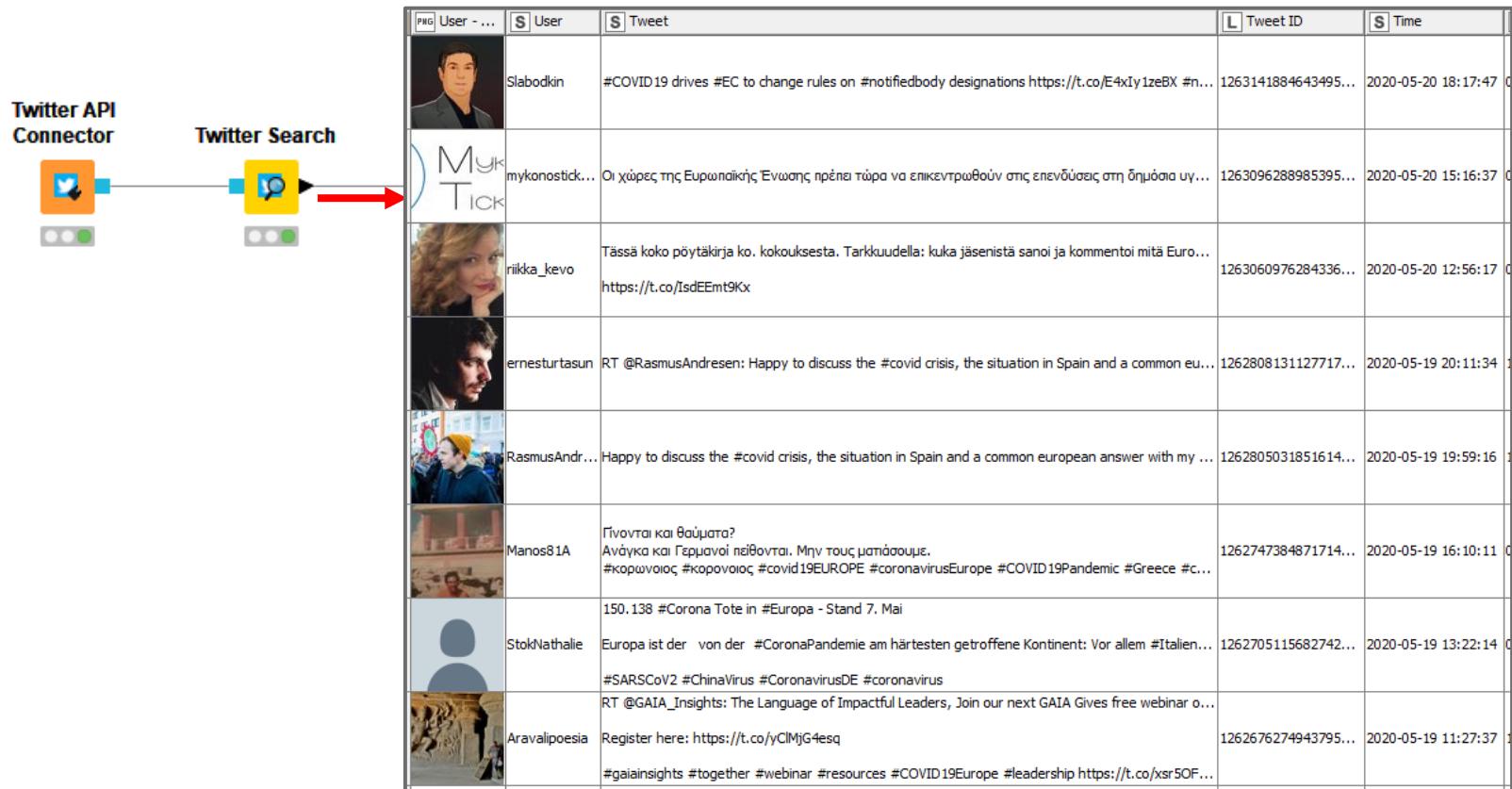
- To import tweets data directly into KNIME via official [Twitter API](#)
- Different nodes for different [tasks](#)
- Twitter API Connectors needs keys from a [Twitter app](#)



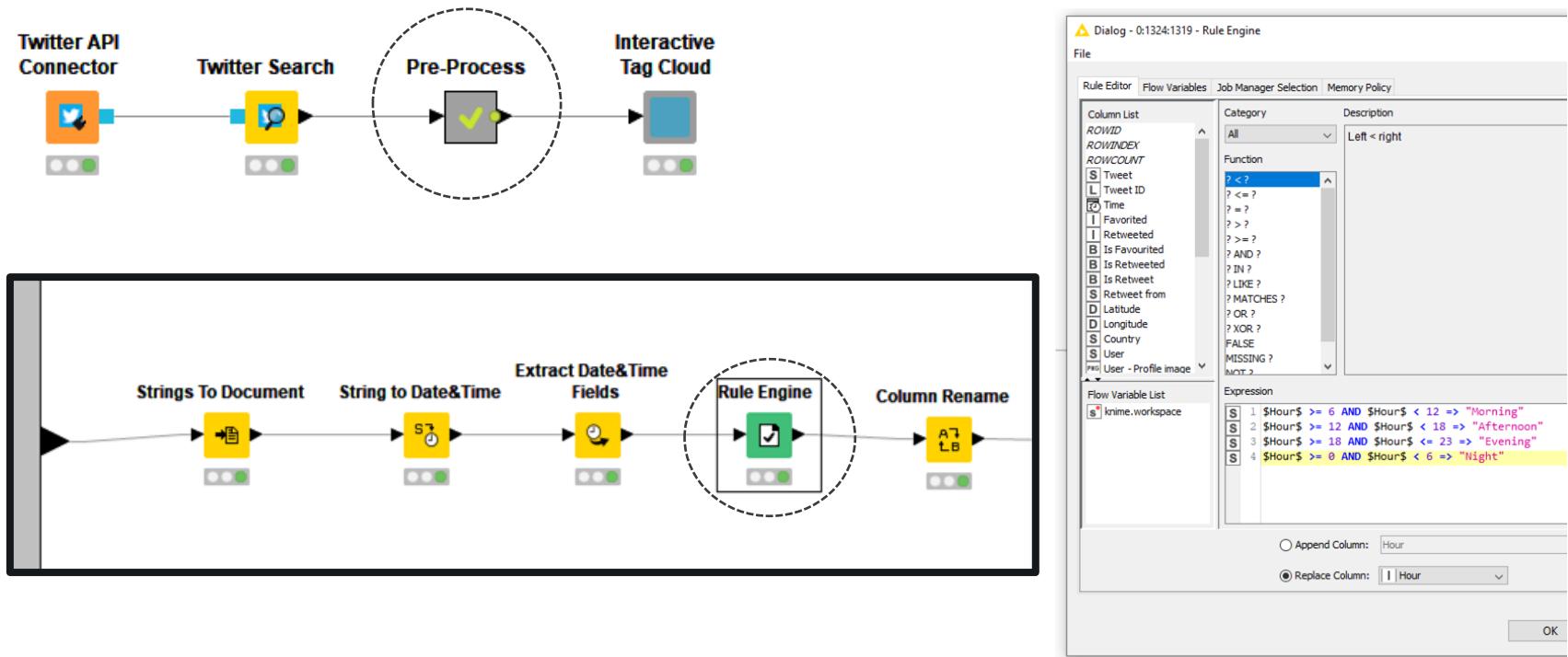
Use Case: Download #COVID19Europe Tweets



Use Case: Download #COVID19Europe Tweets



Use Case: Download #COVID19Europe Tweets



Use Cases



Class Name	Class Abbreviation
Afternoon	Afternoon
Evening	Evening
Morning	Morning
Night	Night

	Colored with Model Labels	User - Profile image
No matching records found		

Showing 0 to 0 of 0 entries (filtered from 990 total entries)

Previous Next

User -
Profile
image

Sentiment Analysis of Social Media Activity



Sentiment Analysis – An Example



Samsung
Samsung Galaxy S7 Edge G935A 32GB Unlocked - Gold Platinum
 125 customer reviews | 606 answered questions

 **Beautiful phone from a wonderful seller!**

By  on May 29, 2017

Color: Gold | **Verified Purchase**

This practically new beautiful phone well exceeded my expectations!



 **One Star**

By  on August 3, 2016

Color: Black Onyx | **Verified Purchase**

Very bad experience



Sentiment Analysis – Some Use Cases

- **Movies:** is this review positive or negative?
- **Products:** what do people think about a new smartphone?
- **Politics:** what do people think about a specific candidate or political issue?
- **Prediction:** predict election outcomes or market trends from sentiment.

Today's Use Case

- Dataset: subset of the IMDb (Internet Movie Database) [Large Movie Review Dataset v1.0](#) with 2000 documents(*).
 - 1000 documents from the positive group
 - 1000 documents from the negative group
- Goal: To assign the correct sentiment label to each document.

(*) For details about the data set see <http://ai.stanford.edu/~amaas/data/sentiment/>

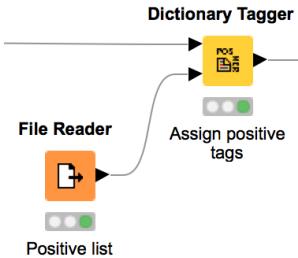
Data citation: Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)

Sentiment Analysis

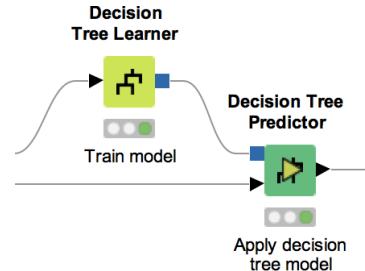
Task: Determine the expressed opinion in a document/text, e.g. positive, negative

Sentiment Analysis = Opinion Mining = Emotion AI

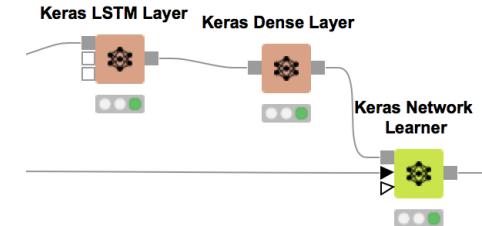
Lexicon Based



Machine Learning



Deep Learning



Approach 1: Lexicon Based

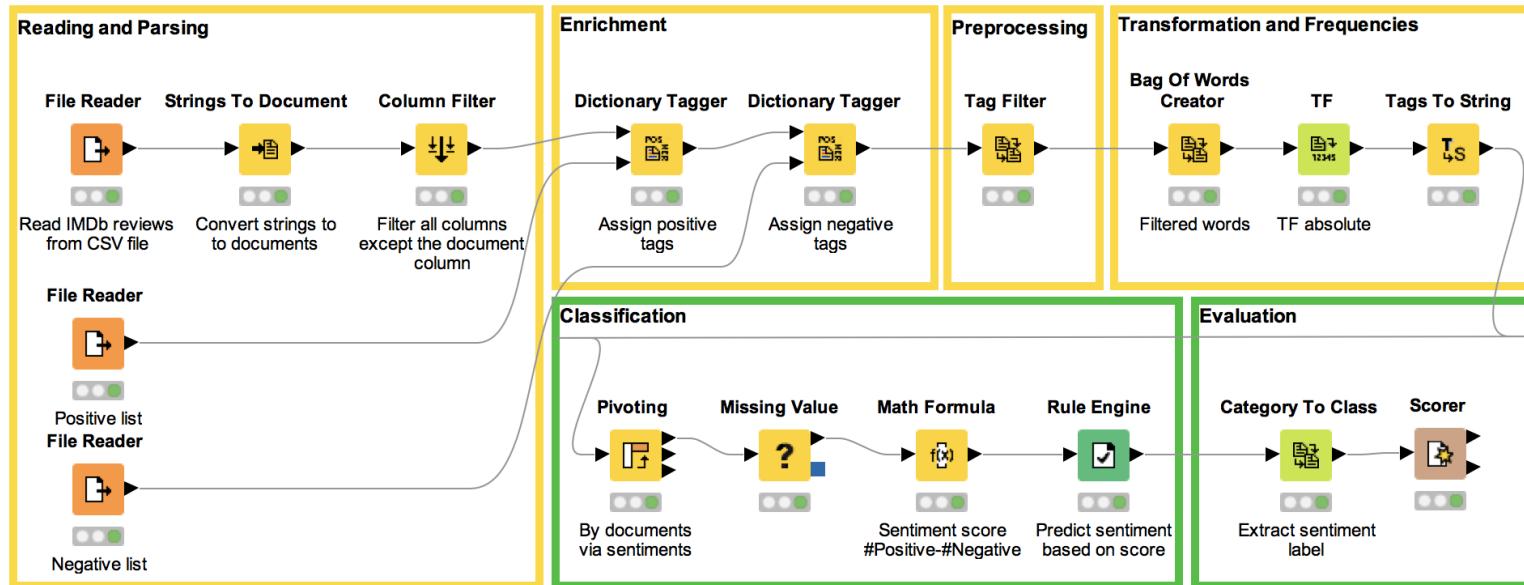
Idea: Rule-based classification with Dictionary Tagger

1. Use custom dictionary to tag positive and negative words
2. Count number of positive and negative words per document
3. Assign classes depending on the number of positive and negative words

$$\text{sentiment score} = \frac{\# \text{ positive words} - \# \text{ negative words}}{\# \text{ total words}}$$

Advantage: No labels needed

Rule-based Classification with Dictionary Tagger



Approach 2: Machine Learning

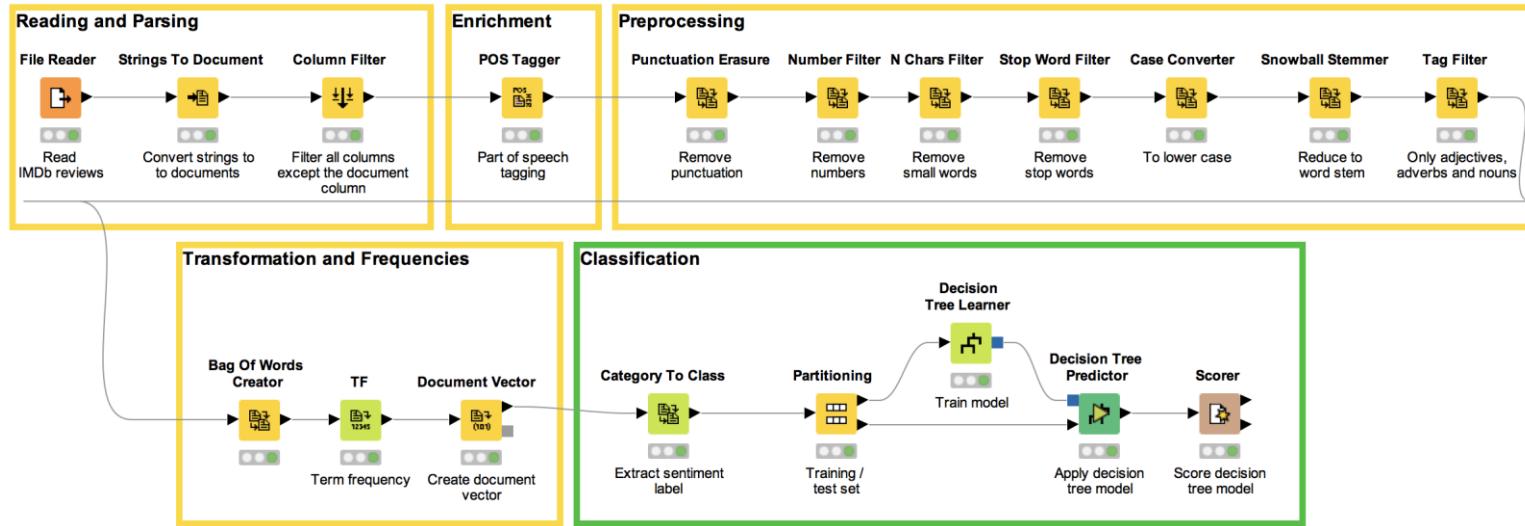
Idea: Train a model to make predictions on unseen data

1. Start the analysis with a labeled dataset
2. Extract a feature space from the documents, e.g. only keywords
3. Train a supervised model, e.g. decision tree, logistic regression

Advantage: Better performance

Vohra, S. M., and J. B. Teraiya. "A comparative study of sentiment analysis techniques." Journal JIKRCE 2.2 (2013): 313-317.

Approach 2: Machine Learning



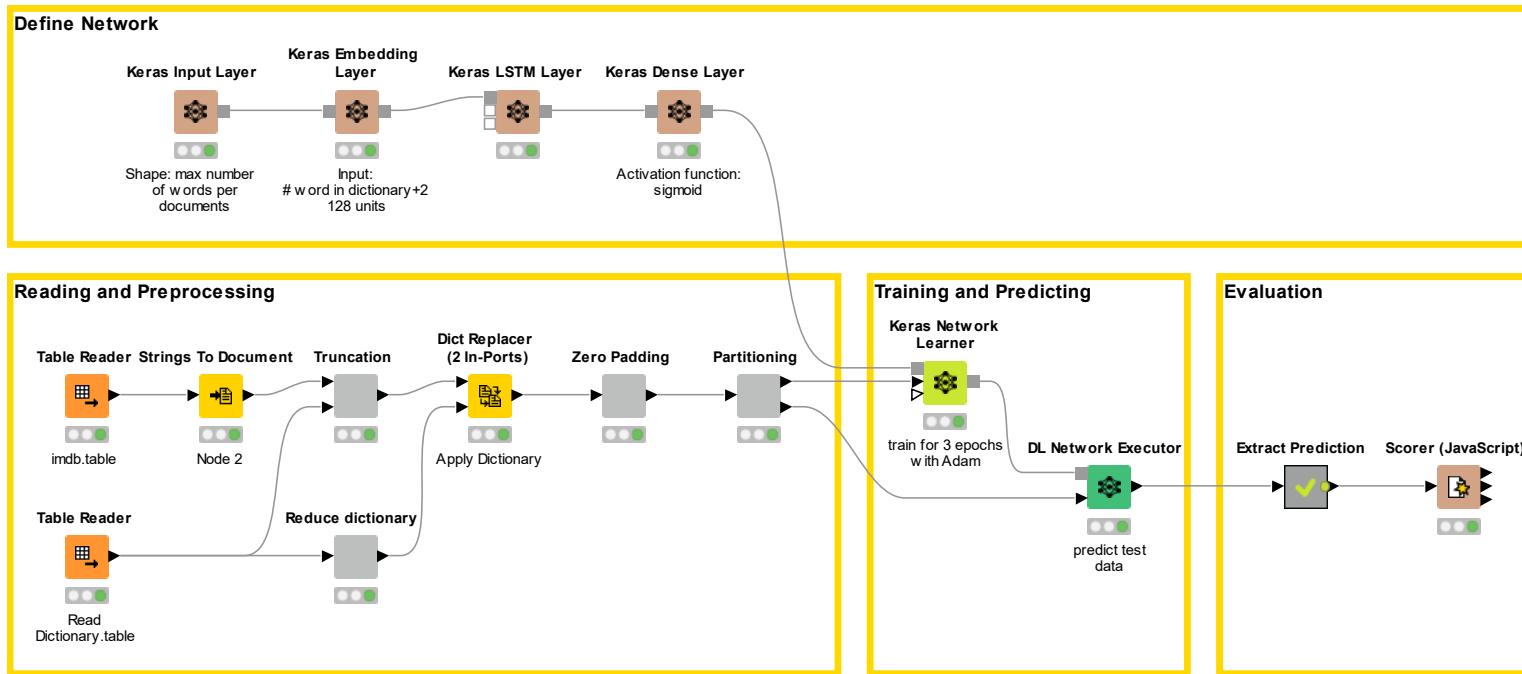
Approach 3: Deep Learning

Idea: Train a network to make predictions on unseen data

1. Start the analysis with a labeled dataset
2. Bring the documents into a standard numeric shape
 - Hot encoding vs. embeddings
 - Zero padding and truncation
3. Define the network structure. Commonly used layers for text are embedding and LSTM layers
4. Train and apply the network

Advantage: Can lead to better performance on big data sets

Approach 3: Deep Learning



Topic Detection in Documents with LDA



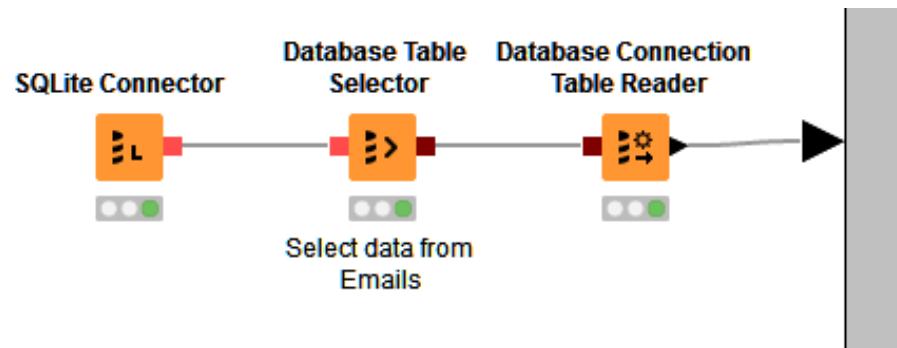
Background Story

- Controversy regarding former Secretary Hillary Clinton's use of a non-government, privately maintained email server
- On 2015, Hillary Clinton released 7,945 emails in response to a U.S. Department of State Freedom of Information Act (FOIA) request
- The original emails are available on the State Department's Website

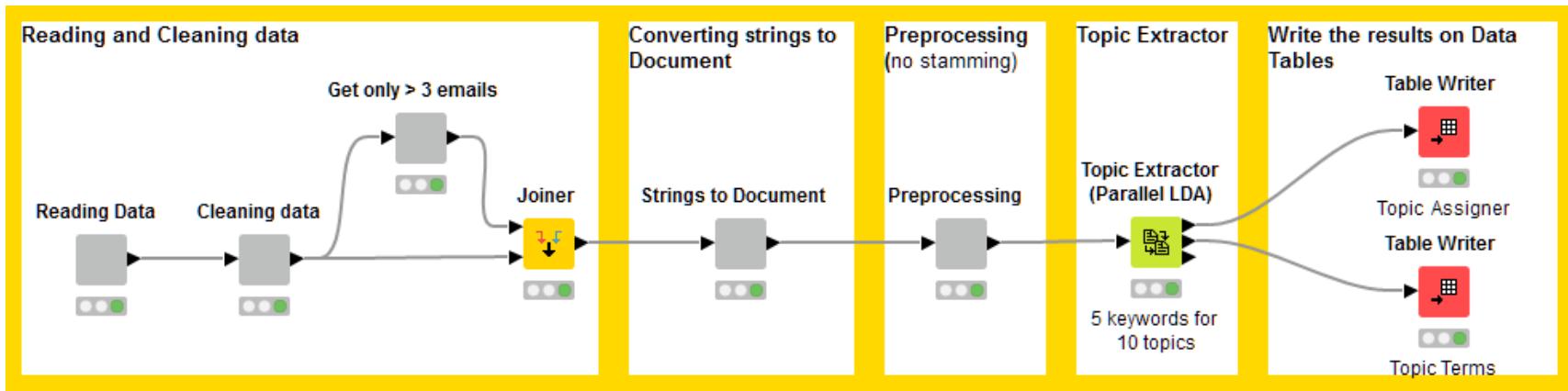


Acquiring and Accessing the Data

- Data available at
kaggle.com/kaggle/hillary-clinton-emails
- The data have been downloaded and accessed as
SQLite database

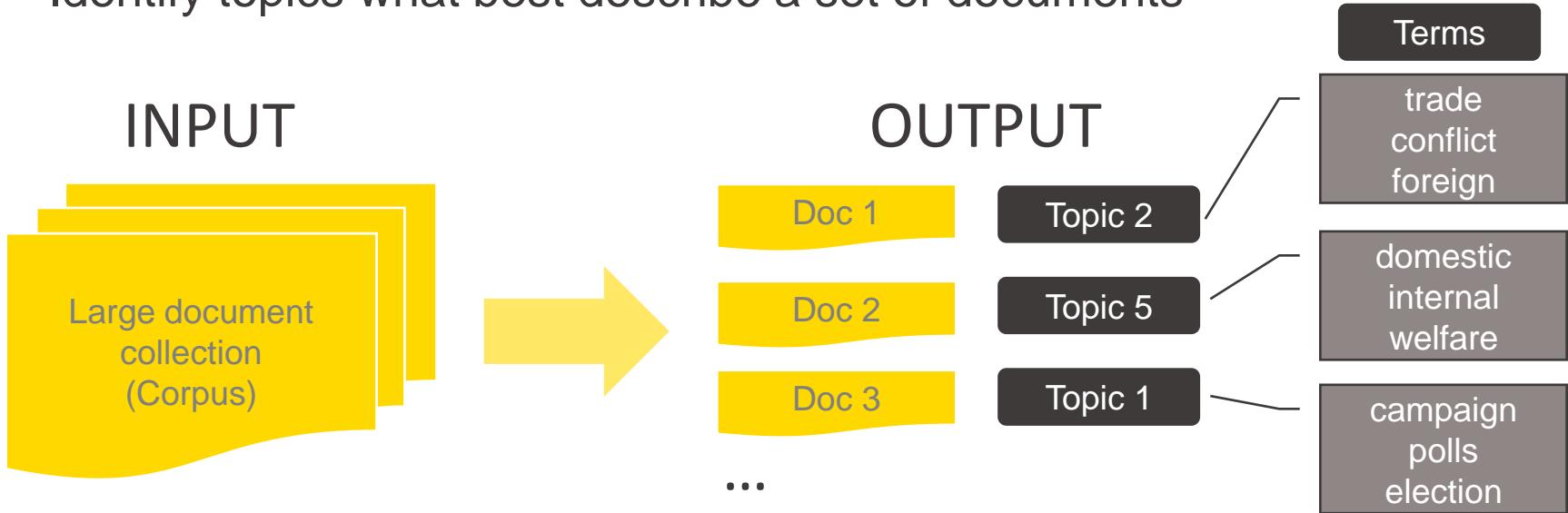


Training Model



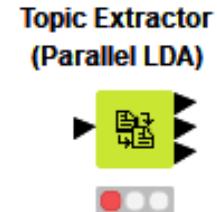
What is Topic Modeling?

- Developed to organize large collections of documents and to **extract the hidden thematic structure**
- Identify topics what best describe a set of documents



Latent Dirichlet Allocation (LDA)

- Automatically finds the top K topics with the most relevant N keywords discussed in a collection of unlabeled documents (considered unsupervised)
- It represents documents as random mixtures over latent topics, where each topic is characterized by a distribution over words
- Syntax or order of the words in document is not important (bag of words model)
- Document order is not important
- The same word can belong to different topics
- **The number of topics needs to be selected/known in advance**

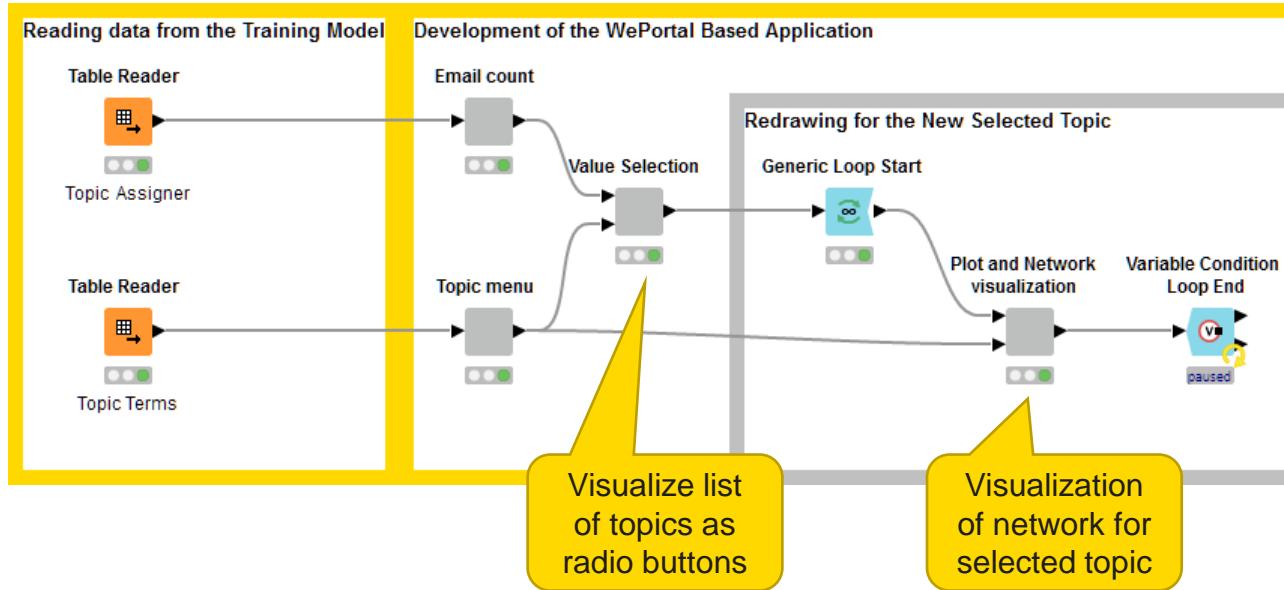


Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993–1022.

Results: Topic and Terms

Topic id	Term
topic_0	[call, tomorrow, talk, fyi, speech]
topic_1	[office, secretary, meeting, department, arrive]
topic_2	[obama, president, house, clinton, white]
topic_3	[china, party, minister, election, labour]
topic_4	[percent, palau, favorable, unfavorable, obama]
topic_5	[benghazi, information, subject, department, doc]
topic_6	[gov, fyi, cheryl, sid, mills]
topic_7	[israel, israeli, afghanistan, military, iran]
topic_8	[party, tea, koch, political, movement]
topic_9	[women, world, people, united, security]

Deployment Model



WebPortal – First Webpage

Open for Innovation **KNIME** Logout

02_Topic Detection Analysis_Deployment_Web Portal 2017-01-04 00.51.28

TOPIC DETECTION ANALYSIS: THE FACTS ABOUT HILLARY CLINTON'S EMAILS

This workflow shows a topic detection analysis and a network analysis on Hillary Clinton's emails.

Data information:

- Throughout 2015, Hillary Clinton has been embroiled in controversy over the use of personal email accounts on non-government servers during her time as the United States Secretary of State;
- According to Mrs Clinton, she sent or received 62,320 emails during her time as secretary of state. She, or her lawyers, have determined about half of those - 30,490, roughly 55,000 pages, were official and have been turned over to the State Department (<http://www.bbc.com/news/world-us-canada-31806907>). There have been a number of Freedom of Information lawsuits filed over the State Department's failure to fully release the emails sent and received on Clinton's private accounts. On Monday, August 31, the State Department released nearly 7,000 pages of Clinton's heavily redacted emails (its biggest release of emails to date);
- The documents were released by the State Department as PDFs. Then the released documents have been cleaned and normalized by some users and are hosting for public analysis. Kaggle's choice to host this dataset (<https://www.kaggle.com/kaggle/hillary-clinton-emails>).

The workflow reads textual data from a SQLite Connector node and converts the strings into documents. The documents are then preprocessed, i.e. filtered and transformed into numerical document vectors. All the preprocessing takes place in the Preprocessing meta node. Then, a topic extraction has been computed through the Topic Extractor (Parallel LDA) node. Finally a loop has been created in order to visualize the data on the KNIME Server.

Please select the topics you are interested in by using the dropdown menu. Then, in order to visualize the Email Frequencies Overtime and the related Network Image click on the Next button. To reload the Plot and the Network Image, please select the topics you are interested in.

Label

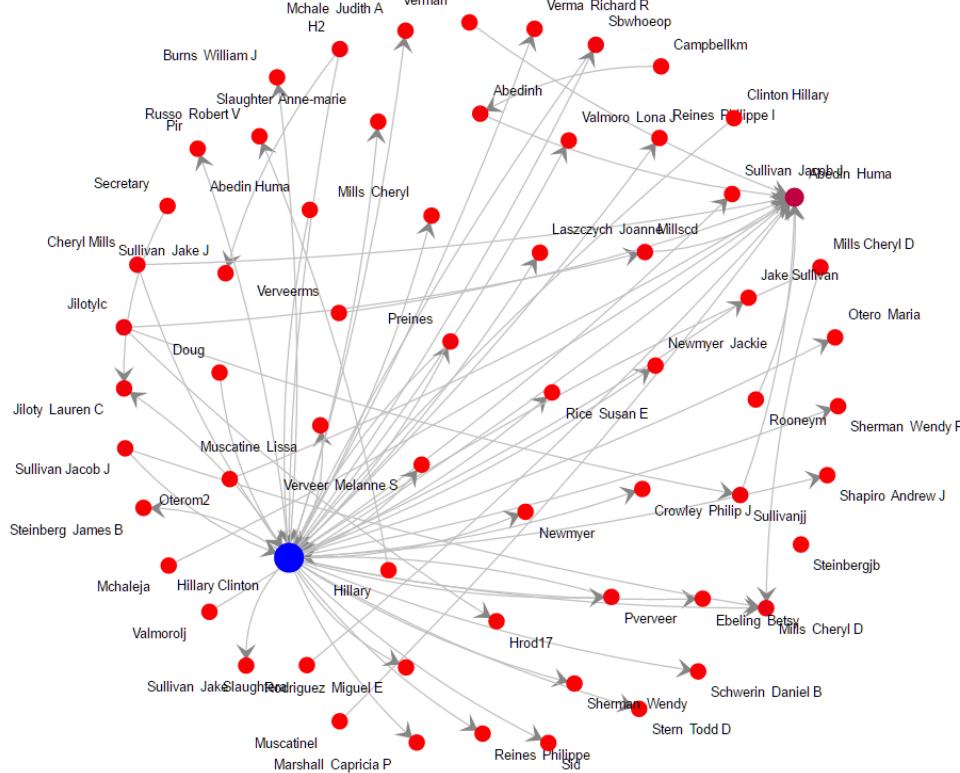
- call, tomorrow, talk, fyi, speech
- office, secretary, meeting, department, arrive
- obama, president, house, clinton, white
- china, party, minister, election, labour
- percent, palau, favorable, unfavorable, obama
- benghazi, information, subject, department, doc
- gov, fyi, cheryl, sid, mills
- israel, israeli, afghanistan, military, iran
- party, tea, koch, political, movement
- women, world, people, united, security
- Exit

Label

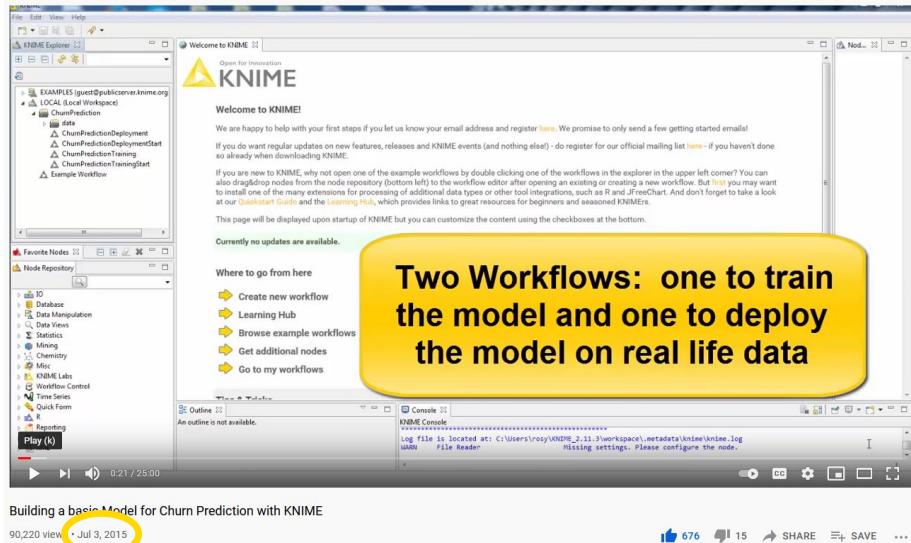
- call, tomorrow, talk, fyi, speech
- office, secretary, meeting, department, arrive
- obama, president, house, clinton, white
- china, party, minister, election, labour
- percent, palau, favorable, unfavorable, obama
- benghazi, information, subject, department, doc
- gov, fyi, cheryl, sid, mills
- israel, israeli, afghanistan, military, iran
- party, tea, koch, political, movement
- women, world, people, united, security
- Exit

[Back](#) [Next >](#)

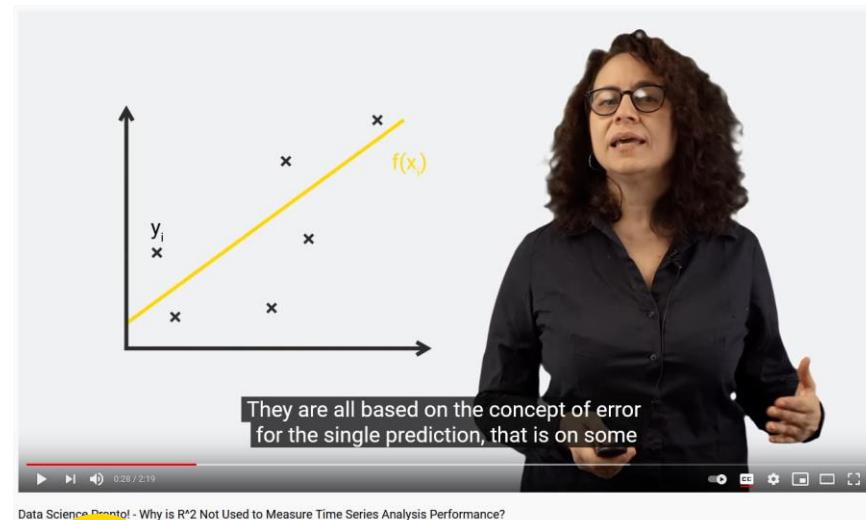
Women, World, People, United, Security



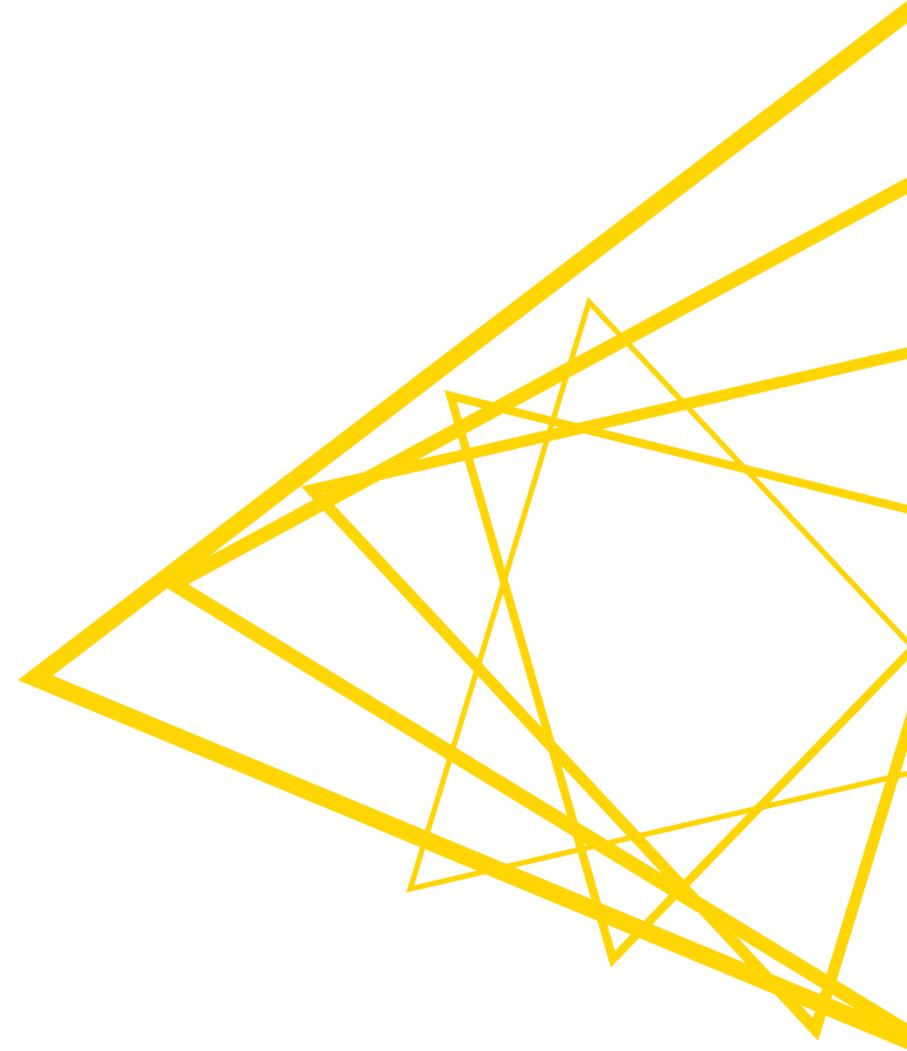
Lesson Learned



- Do not wait till everything is perfect to start ...

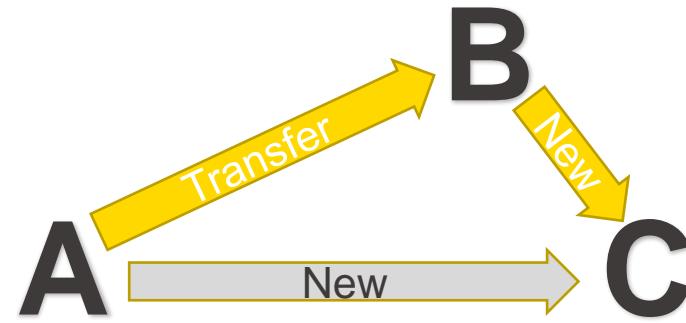


Cancer Cell Classification



What is Transfer Learning?

- Transfer learning can be defined as the attempt to utilize predictive ability in one input / output space to aid in the learning of new input spaces, output spaces, or both.
- In this use case the input space, $3 \times 64 \times 64$, remains the same, but the output space changes from 1000 image categories to 3 cancer types.

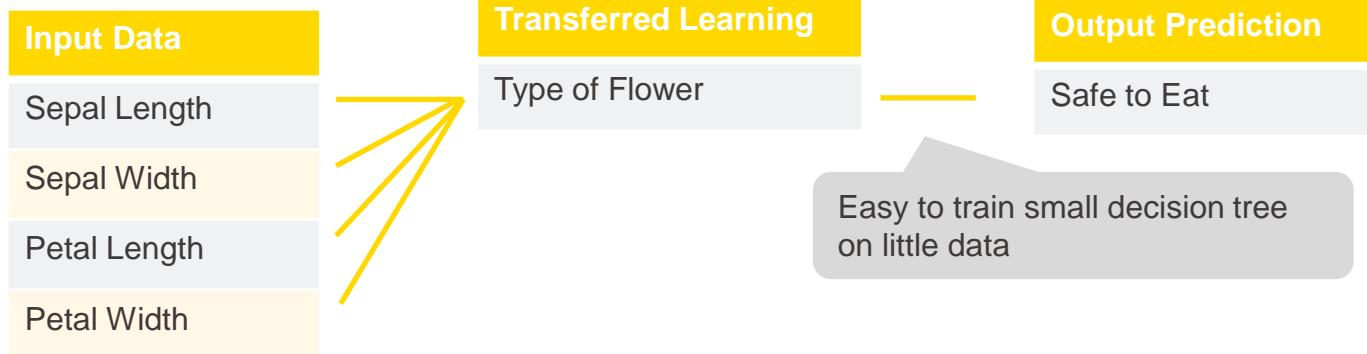


Transfer Learning – Basic Example

Standard Learning

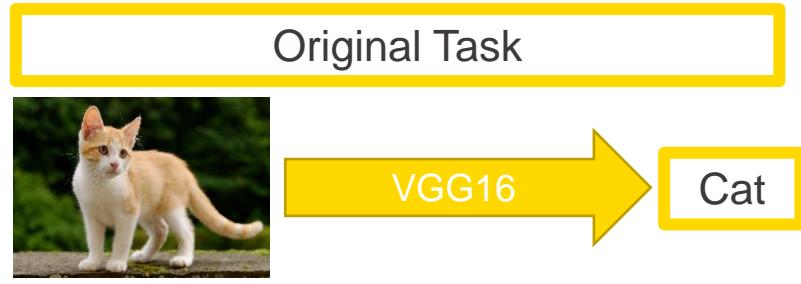


Transfer Learning



Cancer Cell Classification Use Case

- Uses Keras Deep Learning model.
- Learning *Transferred* from VGG16 image classifier
- Completely code free with KNIME's Deep Learning Integration
- Can be adapted to a wide range of image classification problems



[This Photo](#) by Unknown Author is licensed under CC BY-SA

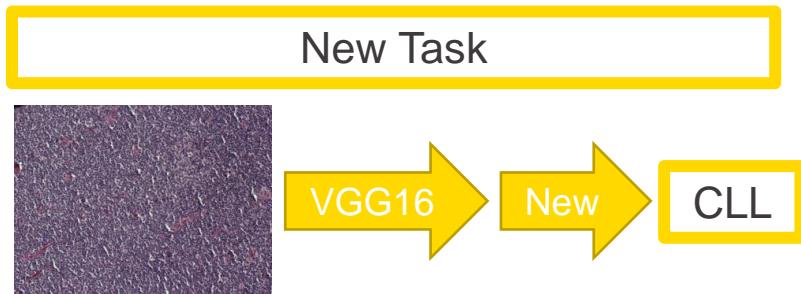
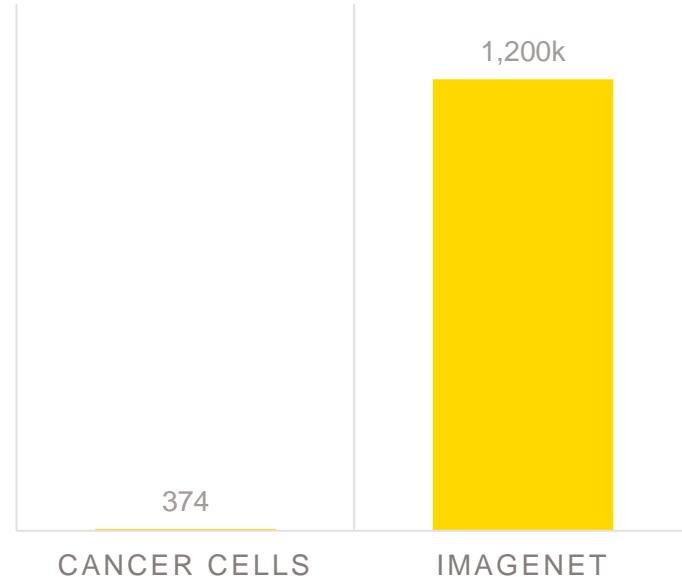


Image From:
<https://ome.grc.nih.gov/licbu2008/lymphoma/index.html>

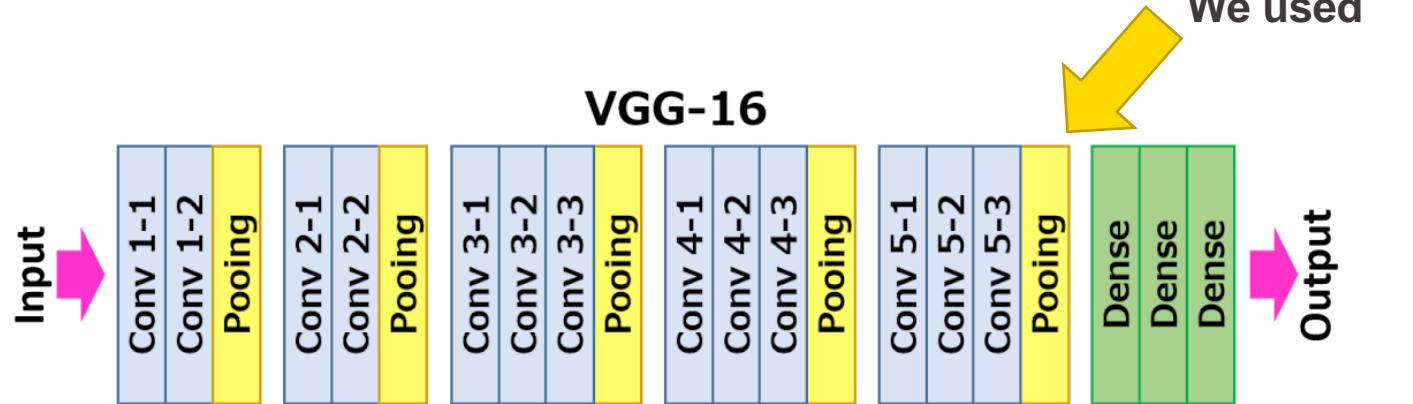
Why is it helpful?

- Deep Learning requires tons of data
- Image classification requires tons and tons of data
- We have 374 labeled cancer images
- VGG16 was trained on more than 1,000,000 images from ImageNet dataset.

DATA POINTS



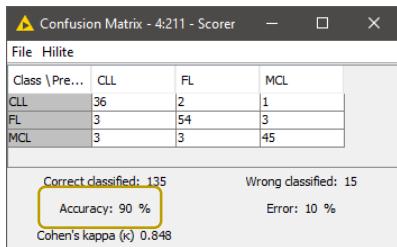
VGG16 Model



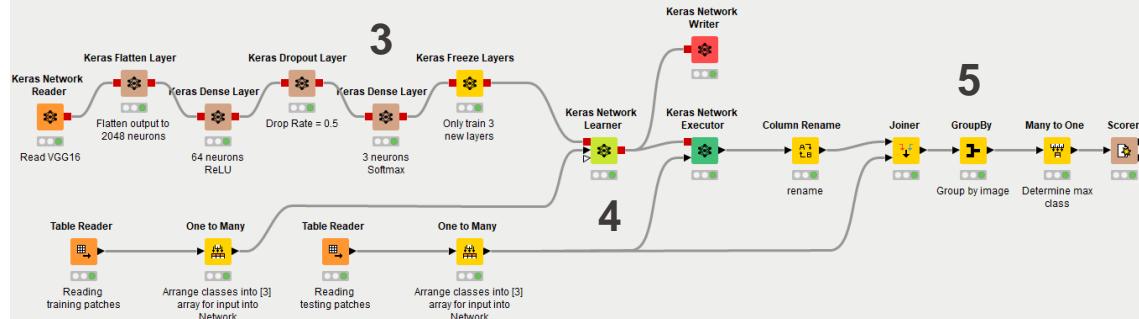
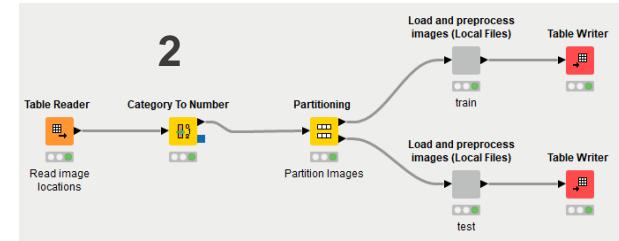
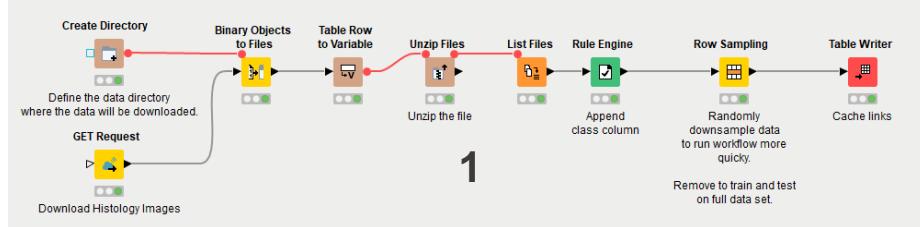
- Several blocks of small convolution layers followed by max pooling layers
- Trained on **ImageNet** dataset
 - Over 1,000,000 images from 1,000 thousand classes
- Won ILSVRC-2013 competition in single network category

Everything in KNIME

1. Download Data Set
2. Pre-Process Data Set
 - Crop, Normalize, Rearrange
3. Configure Network
4. Train Network
5. Post-Process Predictions
 - Group images, create Classification



Workflow on KNIME Hub:
https://kni.me/w/1_w_dxiltpHmsB_QA



Finding the Example

- ✓ EXAMPLES (kname@hub.knime.com)
 - > 00_Components
 - > 01_Data_Access
 - > 02_ETL_Data_Manipulation
 - > 03_Visualization
 - > 04_Analytics
 - > 05_Reportng
 - > 06_Control_Structures
 - > 07_Scripting
 - > 08_Other_Analytics_Types
 - > 09_Enterprise
 - > 10_Big_Data
 - > 20_Strange_but_Educational
 - > 40_Partners
 - ✓ 50_Applications
 - > 02_Credit_Scorng
 - > 03_Churn_Analysis
 - > 04_LastFM_Recommendations
 - > 05_Social_Media_clustering
 - > 07_Forum_Analysis_of_the_KNIME_Forum
 - > 08_RESTDemo
 - > 10_Energy_Usage
 - > 11_Forest_Fire_Prediction
 - > 12_Network_Traffic_Reportng
 - > 13_Address_Deduplication
 - > 14_Medical_Claims
 - > 15_Swiss_Actuarial_Example
 - > 16_MarketBasketAnalysis
 - > 17_Anomaly_Detection
 - > 18_Churn_Prediction
 - > 19_TwitterAnalysis
 - > 21_Model_Selection_and_Management
 - > 24_Customer_Segmentation_Use_Case
 - > 25_DataCleaning_WbPortal
 - > 26_Model_Process_Management
 - > 27_Deployment_Options
 - > 28_Predicting_Departure_Delays
 - > 29_Patent_Network_Analysis
 - > 30_RESTful_ChEMBL
 - ✓ 31_Histopathology_Blog_Post
 - > _data
 - > _legacy_version
 - > _metadata
 - > _metanodes
 - ✗ 00_Keras_Transfer_Learning



Keras Transfer Learning

KNIME Hub > knime > Spaces > Examples > 50_Applications > 31_Histopathology_Blog_Post > 00_Keras_Transfer_Learning

Read Images and Train VGG

Histopathology: Classifying Cancer Cells with Keras
When running this workflow, completely finish executing each section before starting the next. Each portion of the workflow will be automatically refreshed by the next.

1) Download the Dataset: This section automatically downloads the required images for you!

2) Preprocessing: This section reads all the images into a knime format and chops them up into patches that will be fed into our model later.

3) Train Model: the final section reads VGG16, adds layers, trains, and scores our new model.

Download Dataset
This workflow allows the user to download the dataset. Note that the tar.gz file is ~20B.
Original research article:
Meng, Tao, et al. "Histology image classification using supervised classification and multimodal fusion." In: Multimedia (ISM), 2010 IEEE International Symposium on. IEEE, 2010.

Dataset information available at:
<https://omniprcs-rca.nia.nih.gov/Cancer09/Hmpoma/index.html>

Download file from:
<https://omniprcs-rca.nia.nih.gov/Cancer09/Hmpoma.tar.gz>



**Corey**

Open workflow

or download workflow

By downloading the workflow, you agree to our [terms and conditions](#).

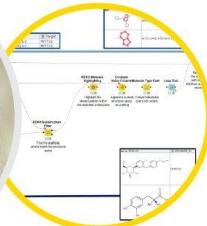
CC-BY-4.0

Short link

https://knl.me/w/1_w_dxiltphMsBQA

Lesson Learned

- Do we need more women in Data Science?
- Is the emergency over?

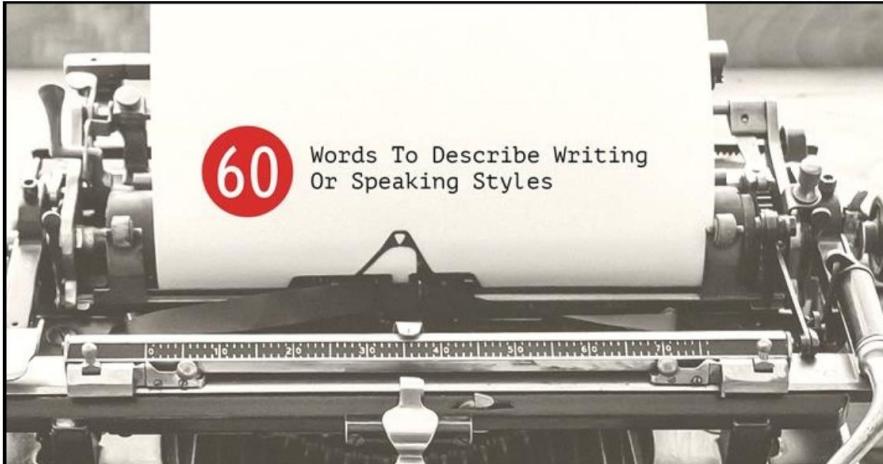


Creative AI: Free Text Generation



Bots & speaking styles

- Articulate
- Chatty
- Clean
- Conversational
- Crisp
- Declamatory
- Diffuse
- Discursive
- Eloquent
- Emphatic
- Epigrammatic
- Epistolary
- Euphemistic
- Flowery
- Funny
- Fluent
- Formal



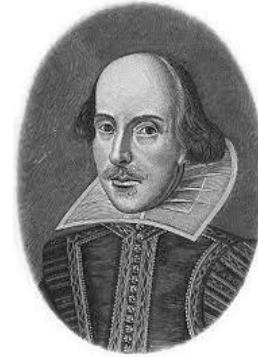
60 Words To Describe Writing Or Speaking Styles

- Gossipy
- Idiomatic
- Incoherent
- Informal
- Journalistic
- Literary
- Lyric
- Ornate
- Parenthetical
- Pejorative
- Picturesque
- Poetic
- Prolix
- Punchy
- Rambling
- Rhetorical
- Rough
- Sesquipedalian

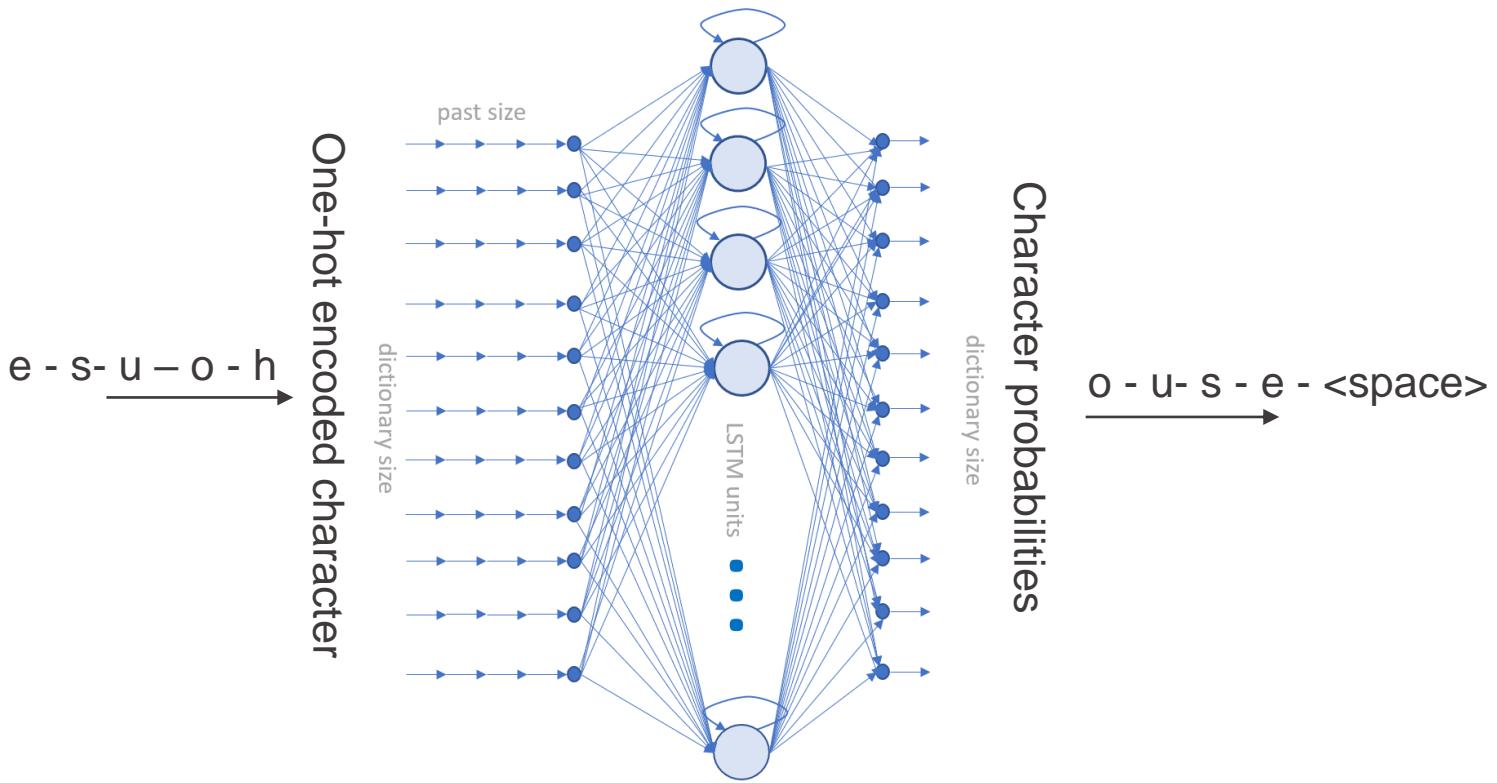
<https://writerswrite.co.za/60-words-used-to-describe-writing-or-speech-style/>

Creative AI: The Problems

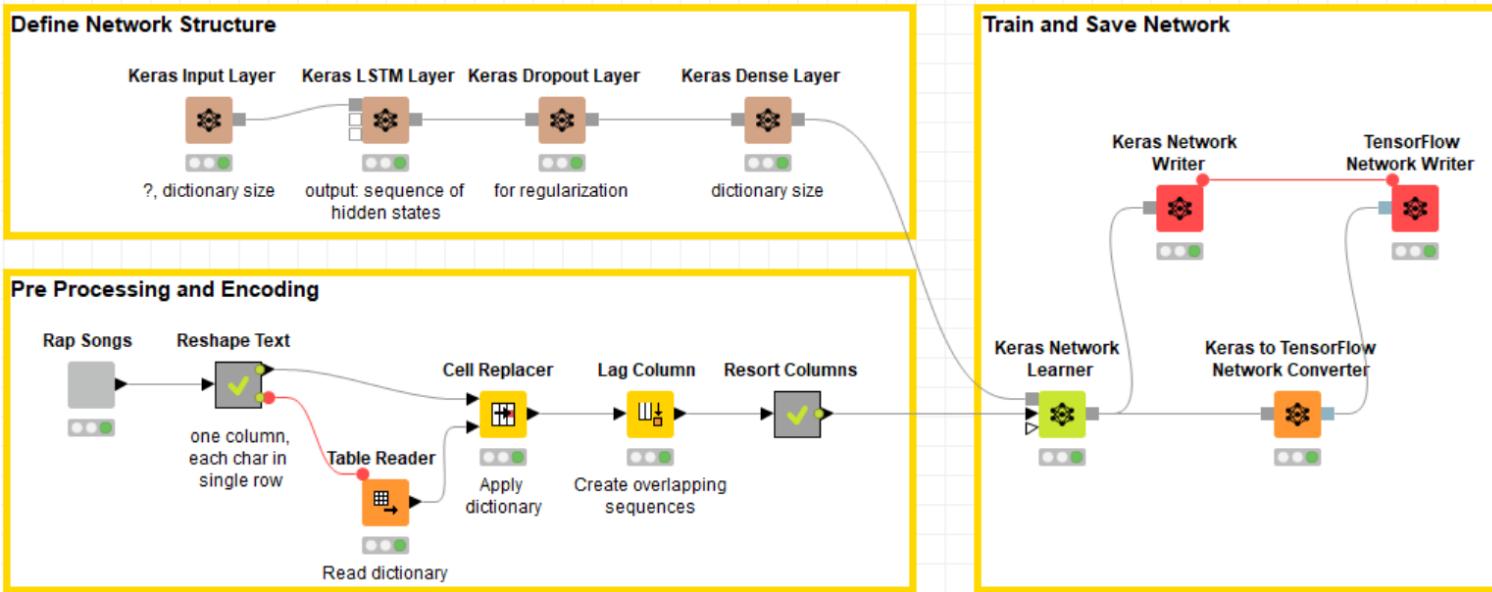
- Free Text Generation
 - Simulating a writing style
 - Writing in different languages
 - Providing an answer in a specific style
 - Generating candidates for product names



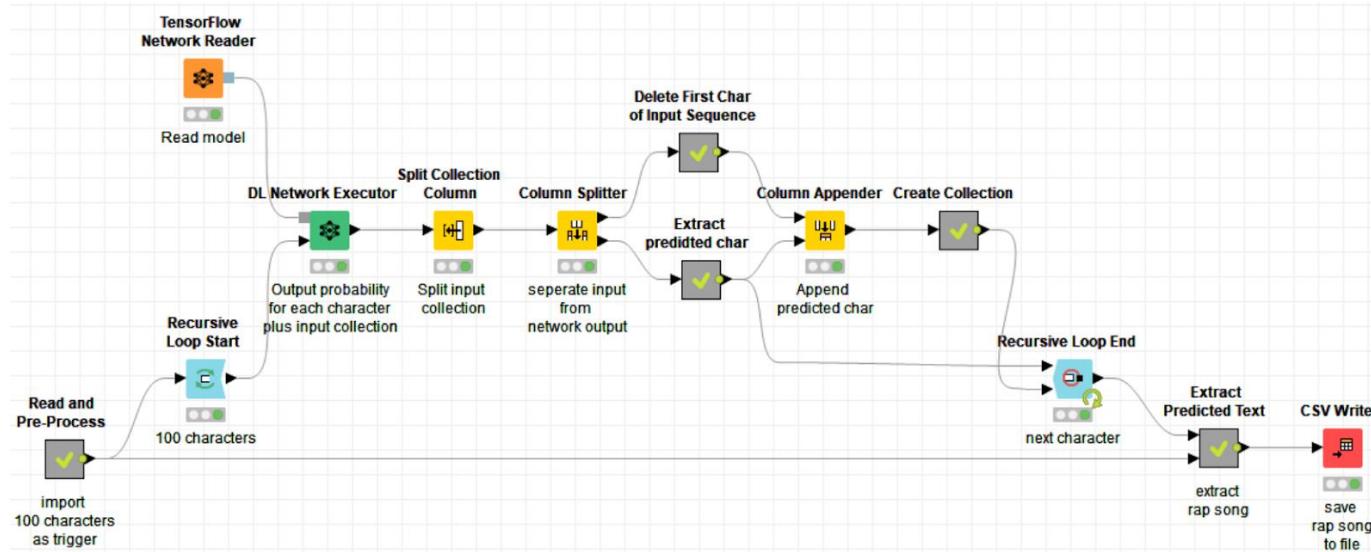
Deep Learning LSTM Network



Creative AI: The Training Workflow

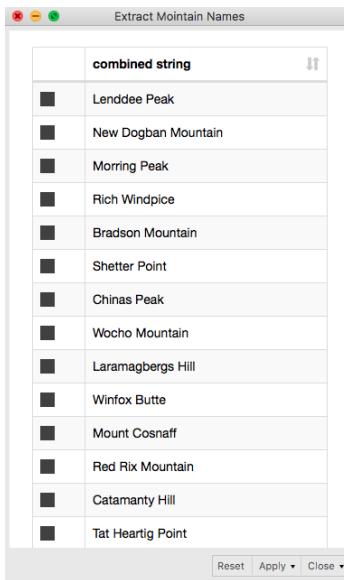


Creative AI: The Deployment Workflow



Creative AI: Deployment and Results

You This post is about generating free text with a deep learning networked the word then I think they don't wanna be the land This that I could take your b**** a** n**** off you in the street or make you were been would 'cause the pigted on my s***, what my tries I need you right in back in a burble It ask me, I can see the winds I swear these n***** a** since when I'm not cool on my for it when I see you n**** that I'm all the one tryna tell me you hot the done of the last this is no other can't expensive, then I got my whip When I get my drow bood and my street with some b**** and the plan is a clear out all my pass that bullshit When I get my dronk If I got my money right it's still the same I'm so high Me and do your hands in the almed and I'm fresher I'm a bad b*** I'm a body in her friend on my whip You ain't with the world the weard and then he give me a n**** dough? And my car is with the bunch of life of the drummers and a crannead and ride with my baby I just lang Mire thes as to see your friend like a m***** shade and started when I came the box I just had to live in a cokin' up in that b***** (b******) You ain't cheap (yeah) Then we feelin' like the president (the phine, watch) If you get shottos (Ahooh) I kick the right things if I said, I get the streets still gon' shoot me so cold weave and make me wanna show me at the top of the back are I want to get me here I can see you a mastac (lean, fuck me fersin') I don't even cold Ten (bottom) I know I gotta keep my shere in a migat (Woo, woo,



This License refers to version of the GNU General Public License. Copyright also means copyright-bick, Remade me any thing to his sword To his salt and most hidden loose to be so for sings, but not in a libutt of his matter than that shall be sure as will be soldye As master compary, do not live in traitor. Bless thy five wits!

-Kent
O pity!
Sir, where is the patience now,
That this is so far from the sea and some bidings to dismantle
So many folds of save and honest.

-Brabantio
I must not think the Turk of Cassio in the strange metting the cribles
of a charmer be the reviling of libe to say
That I can deceive him to the best advantage,
In her prophetic fairs of a little to presently at your powers; whereof I
thank you, sir.

-Albany
Gloucester, I will prove upancy of his sport and first accuriors and
guard and talking on the white.

-King Lear
Where are the thief?
Thou shalt never have the captains at the letter
To the Moor and thing we have not the better shall be sure as worth
if he be anger—

-Regan
I pray you, have a countend more than think to do a proclaim'd there
of my heart, Hot

The words save, honest, thief, master, traitor, and deceive seem to fit the context. Notice also that the dialogue sprouting from the start text of the license agreement interestingly involves mainly minor, less tragic characters from the plays.

Caro amico ti scrivo così mi distraggo un po'E siccome sei molto lontano più forte ti scriverò. Da quella prima folla strana, che aveva preso il suo nome, e di correre alla casa di don Abbondio, con un viso bene di non poterci andar la casa del padre Cristoforo, e gli disse che s'avvicinava all'uscio, e si mise a sparse di corsa, e di stare a sé, verso la strada di servizio, chiesto le parole che gli andavan dall'altra stanza, e con la sua condizione de' cappuccini, e di consigli ricerche di confidenza delle gride, nel suo passaggio, se non pensava con una certa ripugnanza a casa sua, che andavano a scomparire in un campo di buone ragioni che avevan potuto raccogliere i suoi pensieri, e di sopra non senza interrogare, che la sua avventura aveva fatto predicare, e con la forza d'un fatto come fugitive che aveva preso il suo nome, e di correre alla casa di don Abbondio, con un cappuccino di quella sorte, con un certo sospiro, alzando le sue finestre, e le diede un'occhiata in carrozza. Si vendano a metter nelle mani di chi era stato a sedere sur una strada così fatta con le braccia in

Creative AI: A Touch of Art in Thimble Images



Creative AI: The Problems

- Image Neuro-Styling
 - Picasso
 - Botero
 - Matisse
 - Manet
 - ...

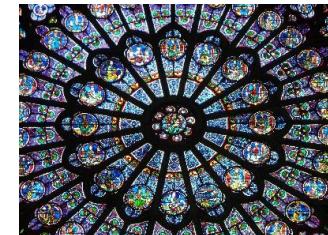


Image Neuro-Styling

Open for Innovation * **KNIME**

Image Input Style transfer (behind the scenes) Results

Neural style transfer

Style your selfie to look like Van Gough Monet painting

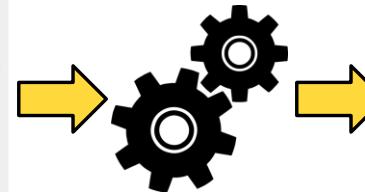
The technique is called Neural Style Transfer and uses pre-trained convolutional neural networks (CNN). The idea is to take content from one image and style from another and blend them together. So as a result you get an image that has the content of the first image but the patterns follow the selected painting. The technique is flexible to learn how to extract style from an arbitrary image and how to apply it via a learned filter.

It is recommended to run this application on a machine with a GPU, as on a CPU the optimisation process runs very slow (up to a couple of minutes), to be compared to several seconds on a modern GPU.

Originality: This technique was suggested in the paper by L. Gatys et al. (2015). This application follows the Keras implementation in [Keras examples](#) and uses VGG19 network pretrained on the ImageNet dataset.

Upload your content image
Please upload the image to be styled.
O'Reilly.jpg

Style image selection
Select images that you want to use to extract style.
Style Images
Medusa



Open for Innovation * **KNIME**

Image Input Style transfer (behind the scenes) Results

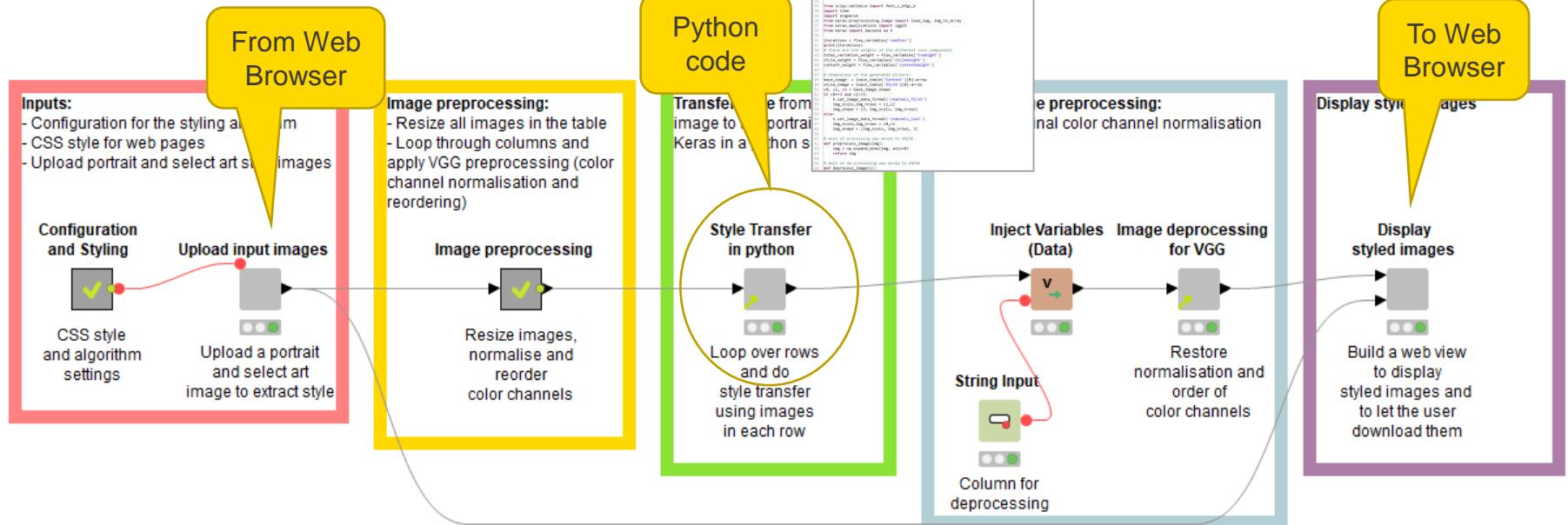
Archive of styled images
Download

Results

Here you can view styled versions of your input image and download a zipped archive with all styled images.

?

Image Neuro-Styling: Workflow



Note, the app requires Deep Learning and python extensions installed. Follow instructions relevant for the Keras extension installation in:
https://docs.knime.com/2018-12/deep_learning_installation_guide/index.html#keras-integration

Rosaria Neuro-Styling



original



Picasso



Magritte



Caravaggio



Notre Dame



Artemisia



Manet



Tapestry



Matisse



Renoir



Van Gogh



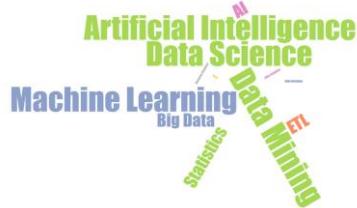
Botero

Lesson Learned

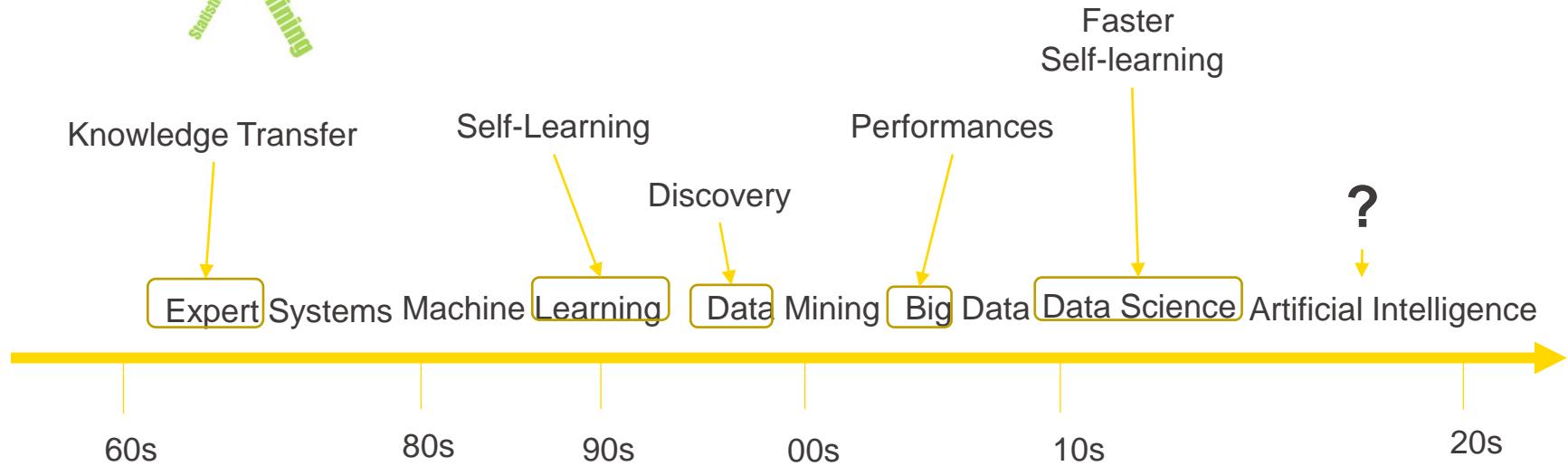
What's in a name? Artificial Intelligence or Data Science?

By Rosaria Silipo Published 10 months ago

3 Comments Like 13 Share Tweet



<https://betanews.com/2019/02/05/artificial-intelligence-or-data-science/>



Free Book as a Thank You

Free Copy of e-book:
“*Practicing Data Science.*
A Collection of Case Studies”

from **KNIME Press**

<https://www.knime.com/knimepress>

with this code: **R-LADIES-0721**

Expires: Sep 30, 2021

