Open for Innovation
KNIME

Welcome to
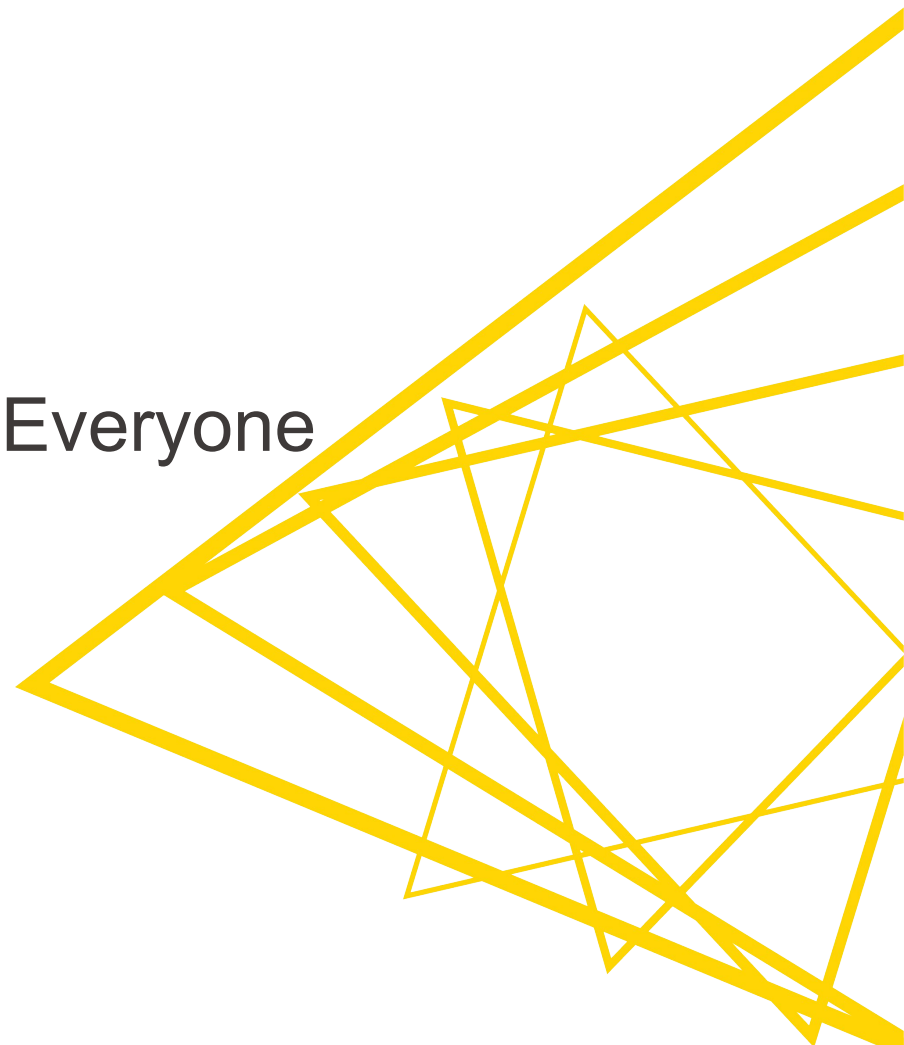# BERT Text Classification for Everyone

Going live at:

Chicago 11:00 am

San Francisco 9:00 am
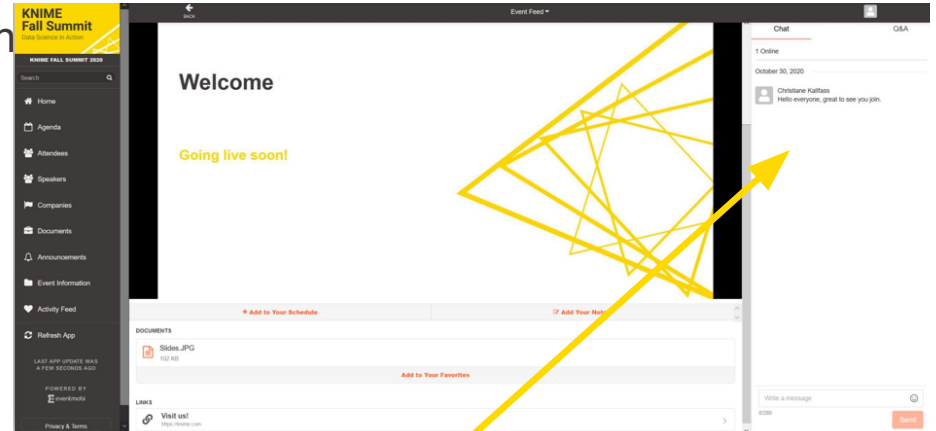
New York 12:00 pm

Berlin 6:00 pm

Housekeeping

- Post in the chat where you are dialing in from and discuss with other attendees

- Questions? Post them in the Q&A

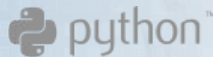Questions will be answered after the presentation.

- **Data Engineering**
- **Data Science**
- **Network analysis**
- **Text mining**
- **Anonymization technologies**

▲ **Rapid application development**

▲ **Reference implementations**

▲ **End to end data science projects**

▲ **Training & support**

▲ **KNIME Server Cloud implementation**

▲ **Node development**

Agenda

- Short introduction to BERT
- BERT nodes for Knime review
- Dataset gathering
- Training parameters settings
- Live demo
- Future developments

Understanding BERT

SQuAD1.1 Leaderboard

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) *Google AI Language* https://arxiv.org/abs/1810.04805	**87.433**	**93.160**
2 Sep 09, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.356	91.202
3 Jul 11, 2018	QANet (ensemble) *Google Brain & CMU*	84.454	90.490

Rank	Model	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	QNLI	RTE
1	BERT: 24-layers, 1024-hidden, 16-heads	80.4	60.5	94.9	85.4/89.3	87.6/86.5	89.3/72.1	86.7	91.1	70.1
2	Singletask Pretrain Transformer	72.8	45.4	91.3	75.7/82.3	82.0/80.0	88.5/70.3	82.1	88.1	56.0
3	BiLSTM+ELMo+Attn	70.5	36.0	90.4	77.9/84.9	75.1/73.3	84.7/64.8	76.4	79.9	56.8

Model	Squad 2.0 test set
ELECTRA-Large	*88.7*
ALBERT-xxlarge	88.1
XLNet-Large	87.9
RoBERTa-Large	86.8
BERT-Large	80.0

(2020)

Source: Google AI Blog

KNIME
Open for Innovation

Understanding BERT

Learn on one task or dataset →	Transfer what has been learnt to another task or dataset
PRE-TRAINING	FINE TUNING

TRANSFER LEARNING

Understanding BERT

PRE-TRAINING

FINE-TUNING
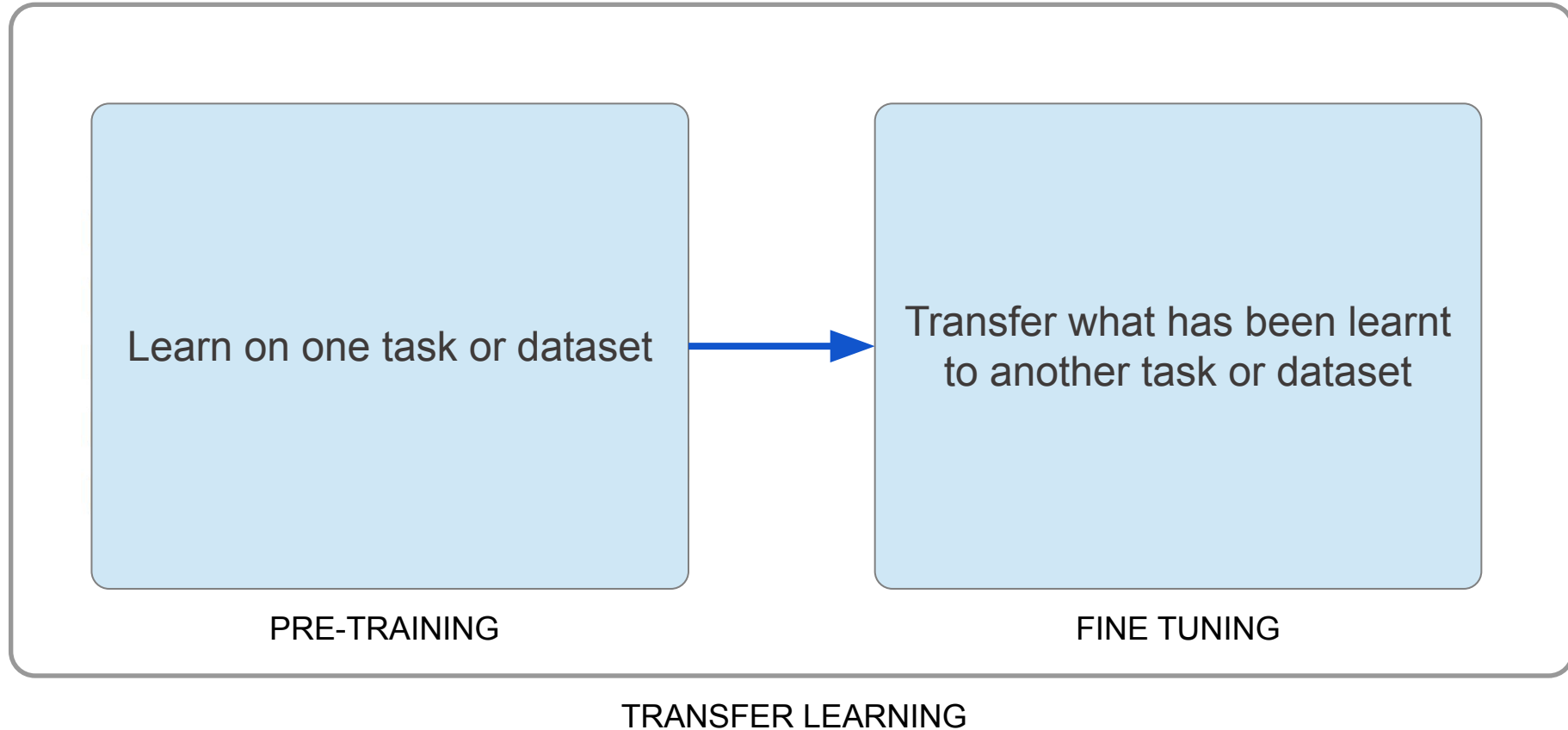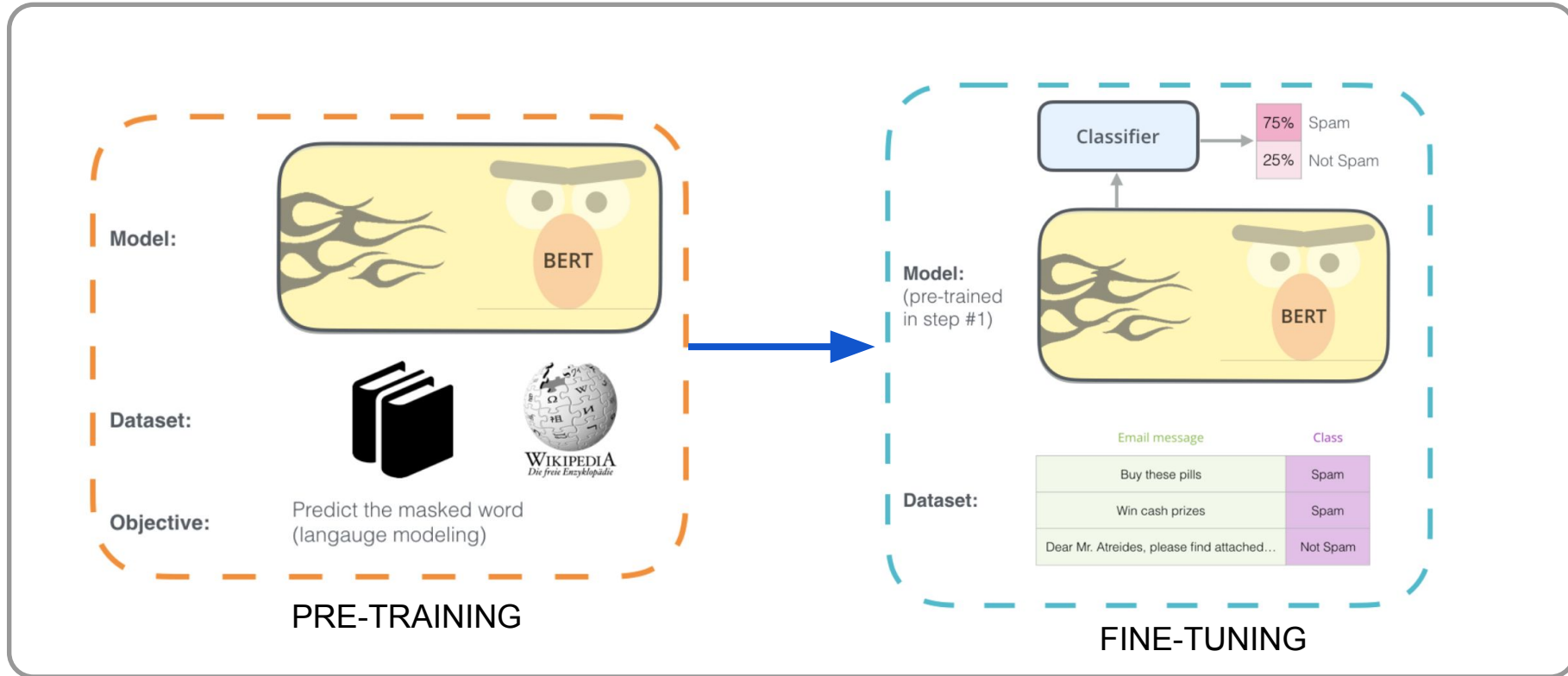
TRANSFER LEARNING

Source: Jay Alammar's Blog

Understanding BERT

- Pre-training task 1: Masked Language Model (MLM)

Input: The man went to the [MASK]$_1$. He bought a [MASK]$_2$ of milk .
Labels: [MASK]$_1$ = store; [MASK]$_2$ = gallon

- Pre-training task 2: Next Sentence Prediction (NSP)

Sentence A = The man went to the store.
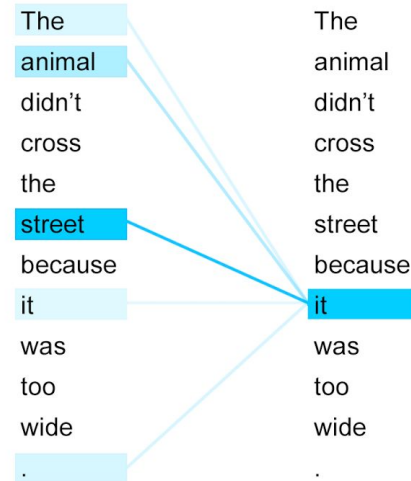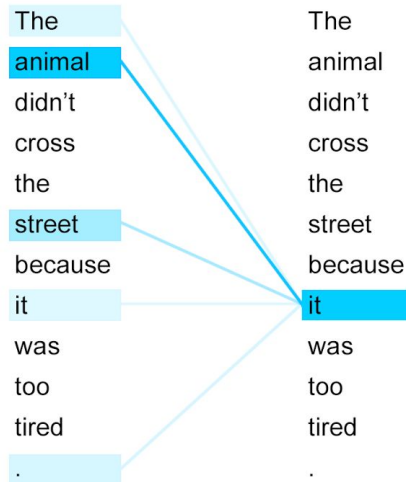**Sentence B** = He bought a gallon of milk.
Label = IsNextSentence

Sentence A = The man went to the store.
Sentence B = Penguins are flightless.
Label = NotNextSentence

KNIME
Open for Innovation

Understanding BERT

- BERT — Bidirectional Encoder Representations from Transformers;
- New neural network nodes called attention;
- Trained on 100+ languages corpuses.

Source: googleblog

BERT extension for Knime

BERT Classification Learner

The node uses BERT model and adds a predefined neural network on top. There are 3 layers added: GlobalAveragePooling1D ...

Community Nodes > BERT by Redfield

Learner

BERT Model Selector

The node allows downloading the model available on TensorFlow Hub. Only models provided by Google are available link . ...

Community Nodes > BERT by Redfield
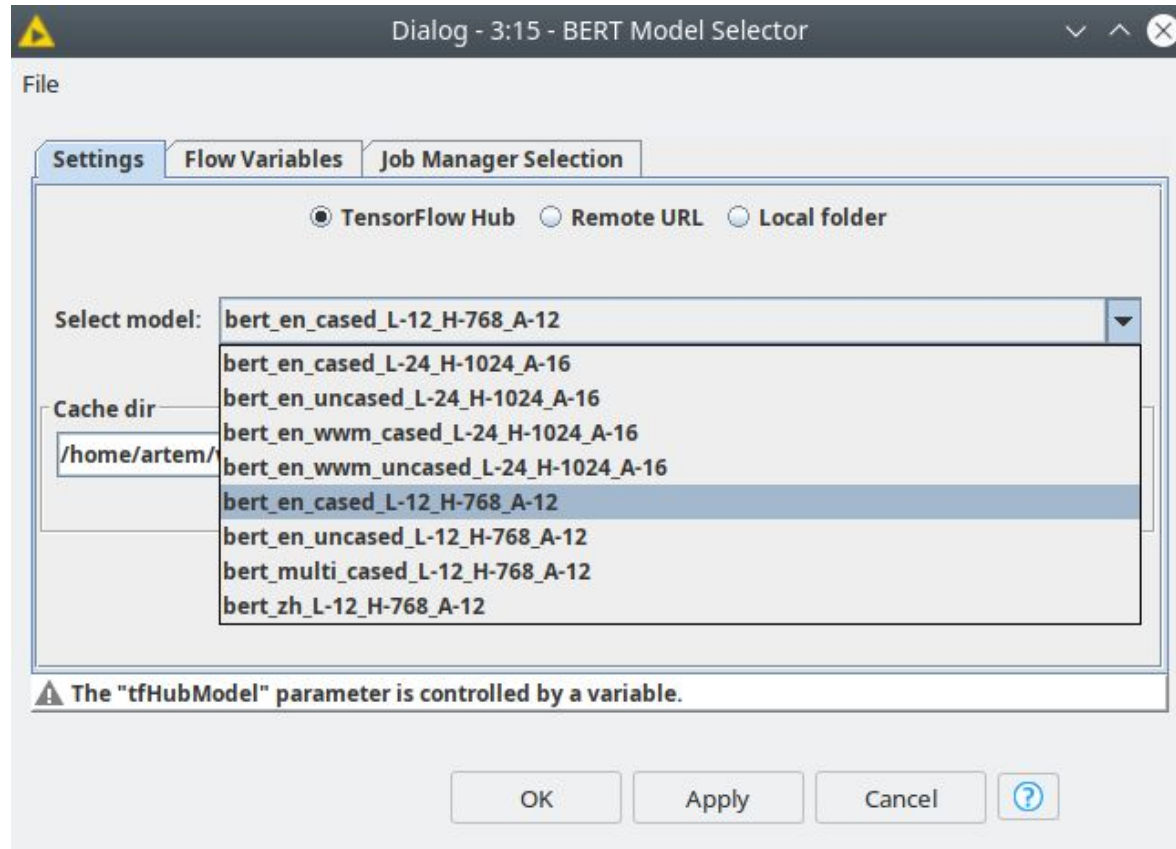
Source

BERT Predictor

The node makes predictions for the input data based on the trained BERT model by BERT Classification Learner.

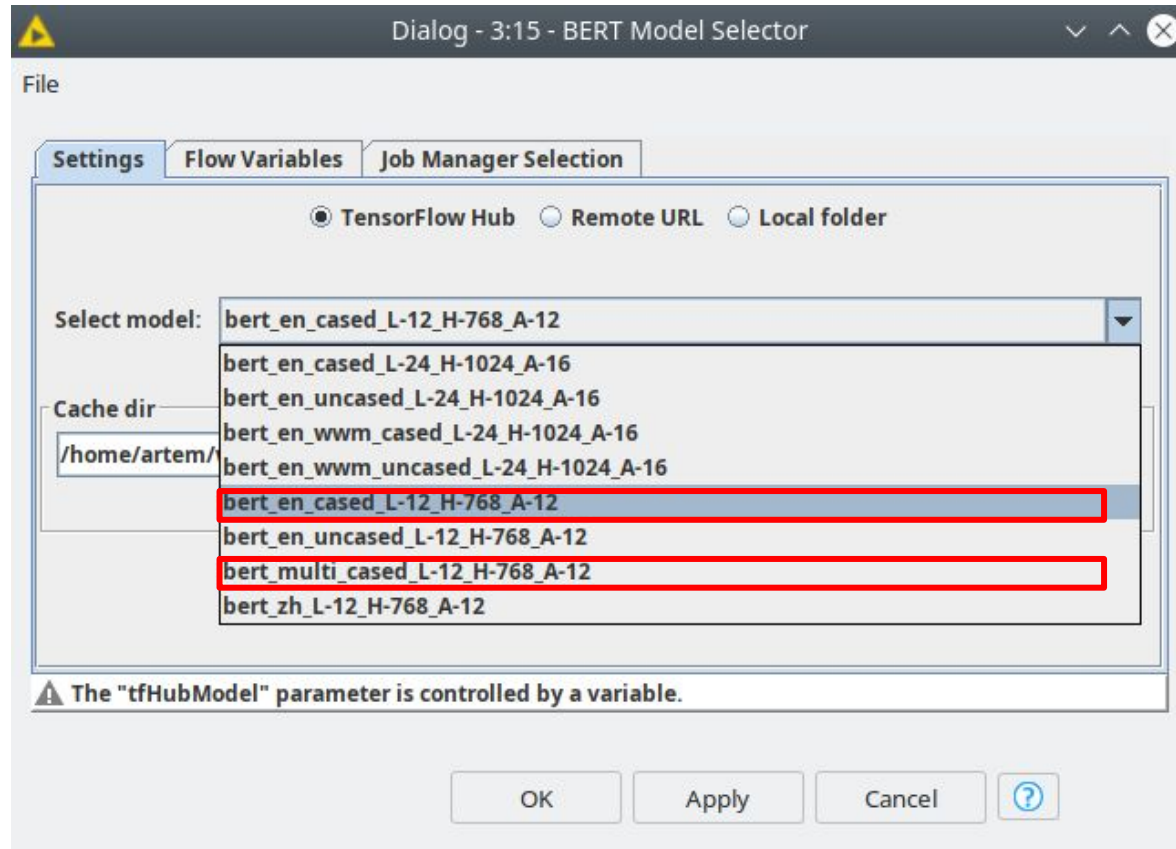Community Nodes > BERT by Redfield

Predictor

TensorFlow hub BERT Models

L — number of transformer blocks; H — hidden layers size; A — number of attention heads.

TensorFlow hub BERT Models

L — number of transformer blocks; H — hidden layers size; A — number of attention heads.

Dataset Gathering and Preparation

- Dataset: https://www.kaggle.com/danofer/dbpedia-classes;
- Enriched with Portuguese texts, ~33K excerpts per language;
- 3 levels of hierarchical labels — 9, 71 and 219 classes per level;

Row ID	wiki_name	text	l1	l2	l3
Row1180...	Dourados_Airport	Francisco de Matos Pereira Airpor...	Place	Infrastructure	Airport
Row1180...	Dublin_Airport	Dublin Airport, (Irish: Aerfort Bhai...	Place	Infrastructure	Airport
Row1180...	Montréal–Mirabel_Int...	Montréal–Mirabel International Ai...	Place	Infrastructure	Airport
Row1180...	Qaqortoq_Heliport	Qaqortoq Heliport (IATA: JJU, ICA...	Place	Infrastructure	Airport
Row1180...	Amsterdam_Airport_S...	Amsterdam Airport Schiphol (Dut...	Place	Infrastructure	Airport
Row1180...	Sisimiut_Airport	Sisimiut Airport (Greenlandic: Mit...	Place	Infrastructure	Airport
Row1180...	Bologna_Guglielmo_...	Bologna Guglielmo Marconi Airp...	Place	Infrastructure	Airport
Row4867...	L'Amour_Toujours_(al...	L'Amour Toujours is the third stu...	Work	MusicalWork	Album
Row4868...	Beg_for_Mercy	Beg for Mercy is the debut album ...	Work	MusicalWork	Album
Row4869...	Make_It_Last_Forever...	Make It Last Forever is the debut ...	Work	MusicalWork	Album
Row4870...	Midget_Tossing	Midget Tossing is the debut albu...	Work	MusicalWork	Album
Row4871...	To_the_Extreme	To the Extreme is the major label ...	Work	MusicalWork	Album
Row4872...	The_Last_Supper:_Liv...	The Last Supper: Live at Hammers...	Work	MusicalWork	Album

Training parameters settings

- The threshold for class occurence is 100;
- Mean text length = 130, median = 96, min = 6, max = 710;
- epochs = 3, batch size = 28, max_seq_length = 128, with BERT fine-tuning;
- Adam optimizer, learning rate = 1E-5.

Training parameters settings

Today's workflow

DBPedia texts classification with BERT by Redfield:
https://kni.me/w/F36Skypp8tyDSHEH

The workflow uses enhanced version that includes Portuguese texts. English data   is taken from here:
https://www.kaggle.com/danofer/dbpedia-classes
The workflow show how to use BERT extension for Knime by Redfield to train models for text classification.
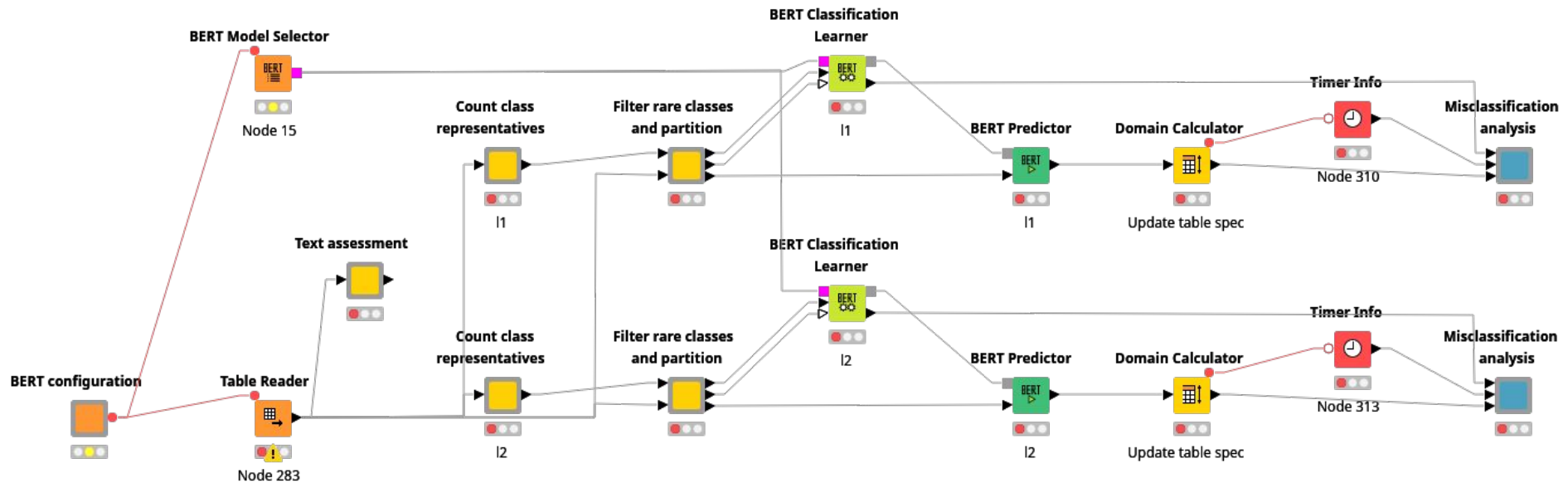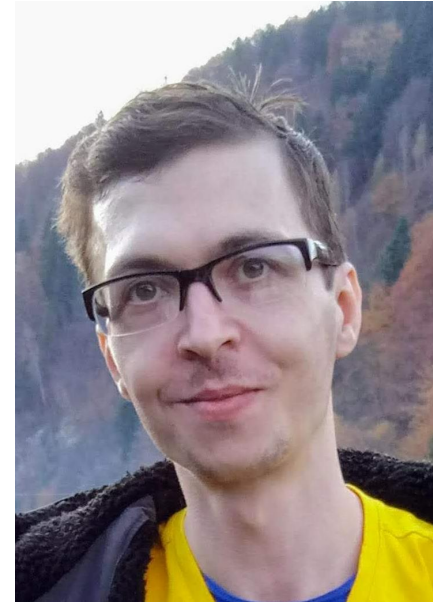
Future Development

- Text classification node able to handle Google TensorFlow Hub models;

- Under development: support of HuggingFace BERT models;

- Other improvements: multi-label classification;

- Other use cases: question answering, similarity, ABSA, NER,…

- Blogpost is coming soon!

Acknowledgements

Nadjet Bouayad-Agha,
PhD and expert in NLP

Alexander Bondaletov,
Java developer

Open for Innovation

KNIME

Thank you for joining!

Artem Ryasik:
artem.ryasik@redfield.se

Redfield AB, Fleminggatan 15A,
112 26 Stockholm
Sweden
info@redfield.se