Custom Translation Using KNIME and Keras

9th October 2018 | Mohammed Ayub and Joseph Gochal | Research and Data Analytics | National Fire Protection Association (NFPA)

NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

NFPA Introduction

Vision - “ We are the leading global advocate for the elimination of death, injury, property and economic loss due to fire, electrical and related hazards”

- Non-profit organization delivering information and knowledge through more than 300+ Code and Standards relating to fire and electrical safety.
- Providing training, education, research and outreach materials to more than 50,000 members around the world
- Research and Data Analytics (RDA) group focusses on building interactive tools and services to enable fire data analytics

 © National Fire Protection Association. All rights reserved.

Top Languages by Total Speakers

Rank	Language	Total Speakers (millions)
1	English	1.12
2	Chinese (incl. Standard Chinese)	1.1
3	Hindi/Urdu	697
4	Spanish	513
5	Arabic	422
6	French	285
7	Malay	281
8	Russian	264
9	Bengali	262
10	Portugese	236

 © National Fire Protection Association. All rights reserved.

NFPA Current State of Affairs

- Outsource most translation efforts, reviewing done by In-house experts
- Focused mainly on Spanish Portuguese and Arabic

Code and Standards

- Colaborative programs like:
  - Latin America Chapters (6)
  - MENA Advisory Committee
  - Academic institutions in Sweden, UK

Global Outreach

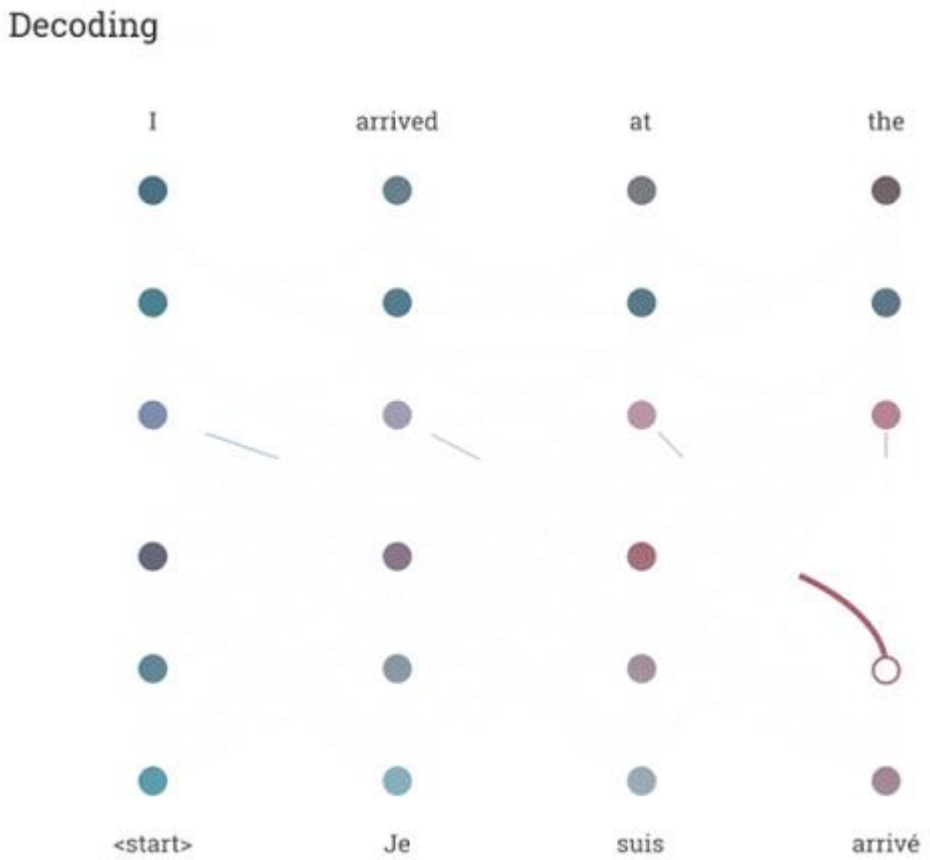
- ~80% Viewership and ~90 Product consumers are form North America Region. Rest is mostly spread across Asia and Middle East.
- NFPA web pages in Spanish

Product Support

 © National Fire Protection Association. All rights reserved.

Neural Machine Translation

- Converting a sequence of symbols in one language to another
- Encoder Decoder architecture
- Variations – RNN's, CNN's and mixed type



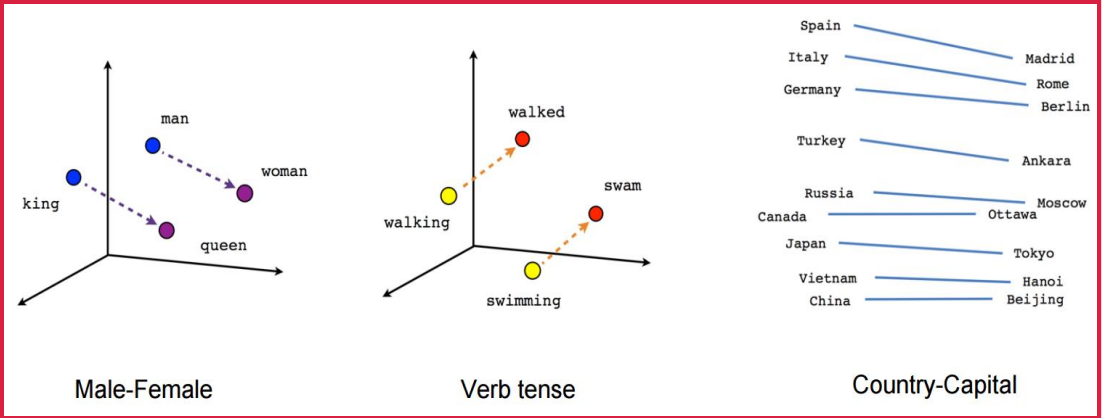
 © National Fire Protection Association. All rights reserved.

Sample:
“ Section 3.2.1.2 Fire-rated glazing assemblies marked as complying with hose stream requirements (H) shall be permitted in applications that do not require compliance with hose stream requirements.”

Preprocess:
“Arc-Fault Circuit Interrupter (AFCI)”
“Circulating Closed-Loop Sprinkler System”

Considerations

Word Embedding:



Sub Words Units – BPE:

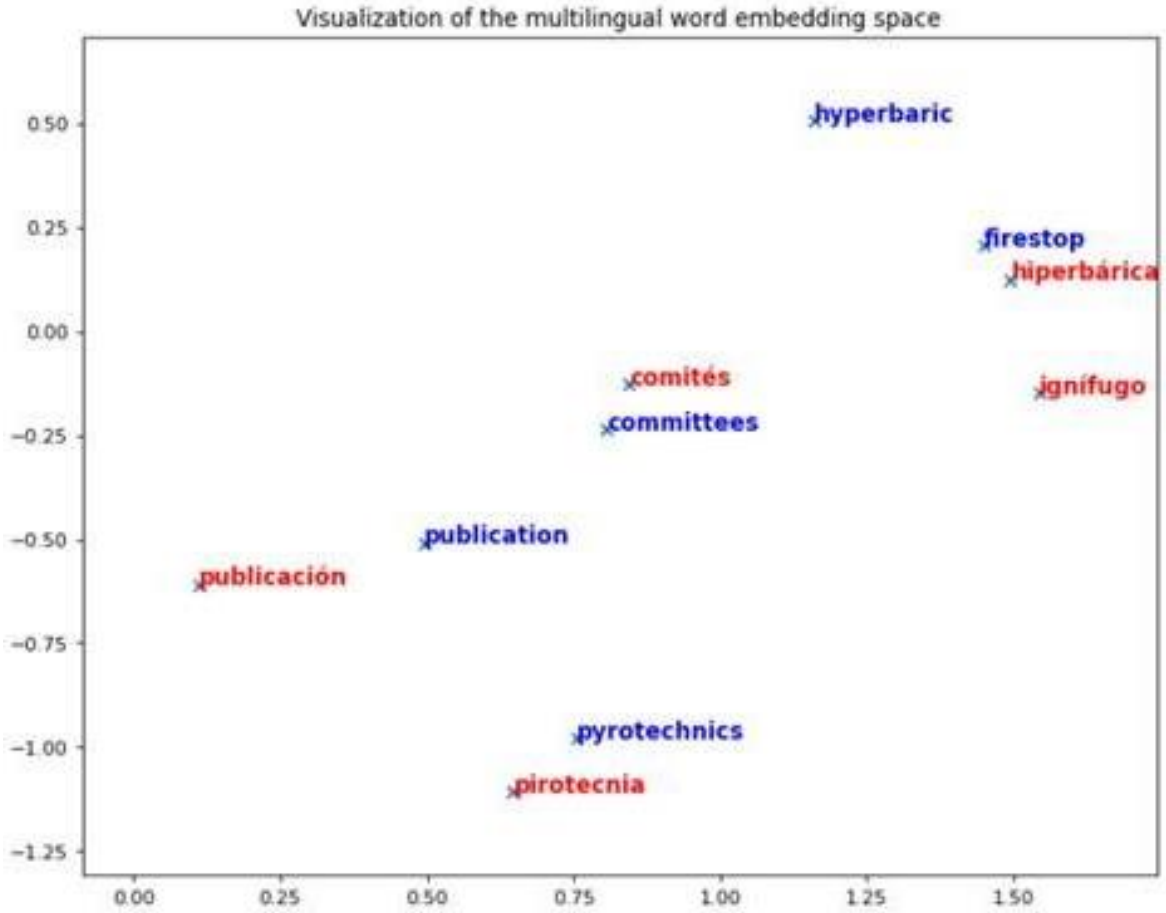
Data Compression Algorithms for Word Segmentation

solar system (English)
Sonnensystem (Sonne + System) (German)
Naprendszer (Nap + Rendszer) (Hungarian)

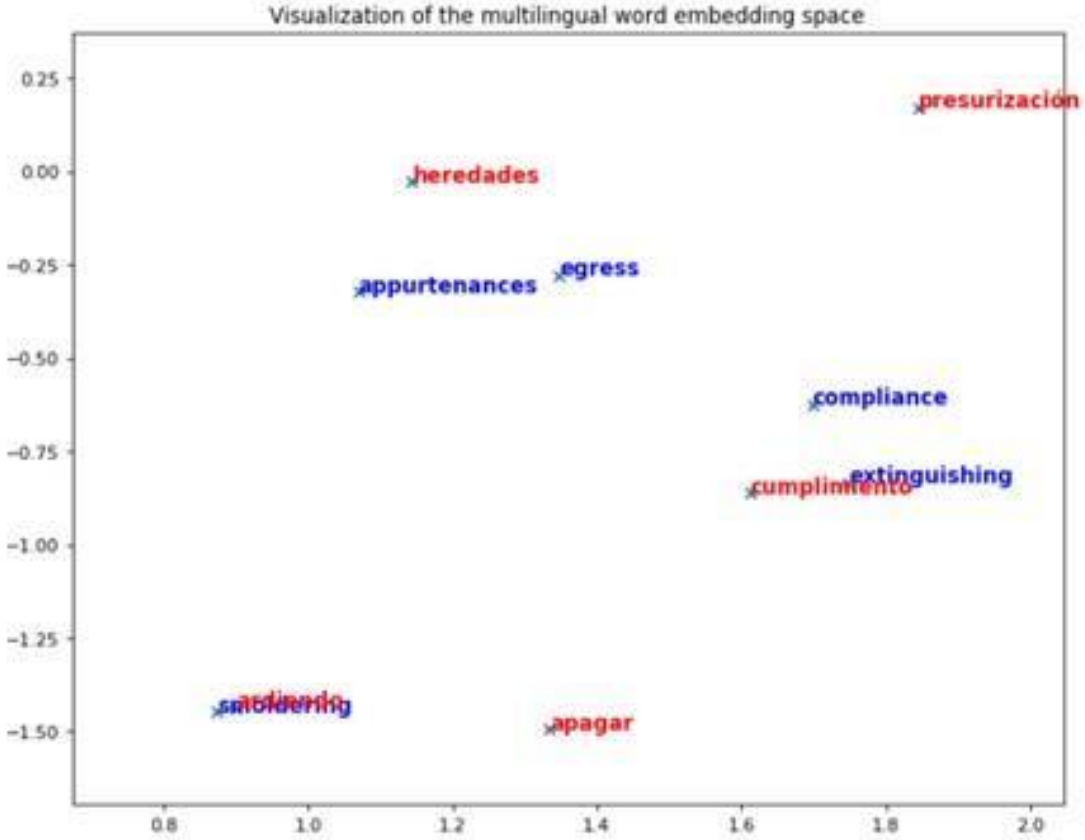
 © National Fire Protection Association. All rights reserved.

Why Domain Adaption is hard ?

Good



Not so good



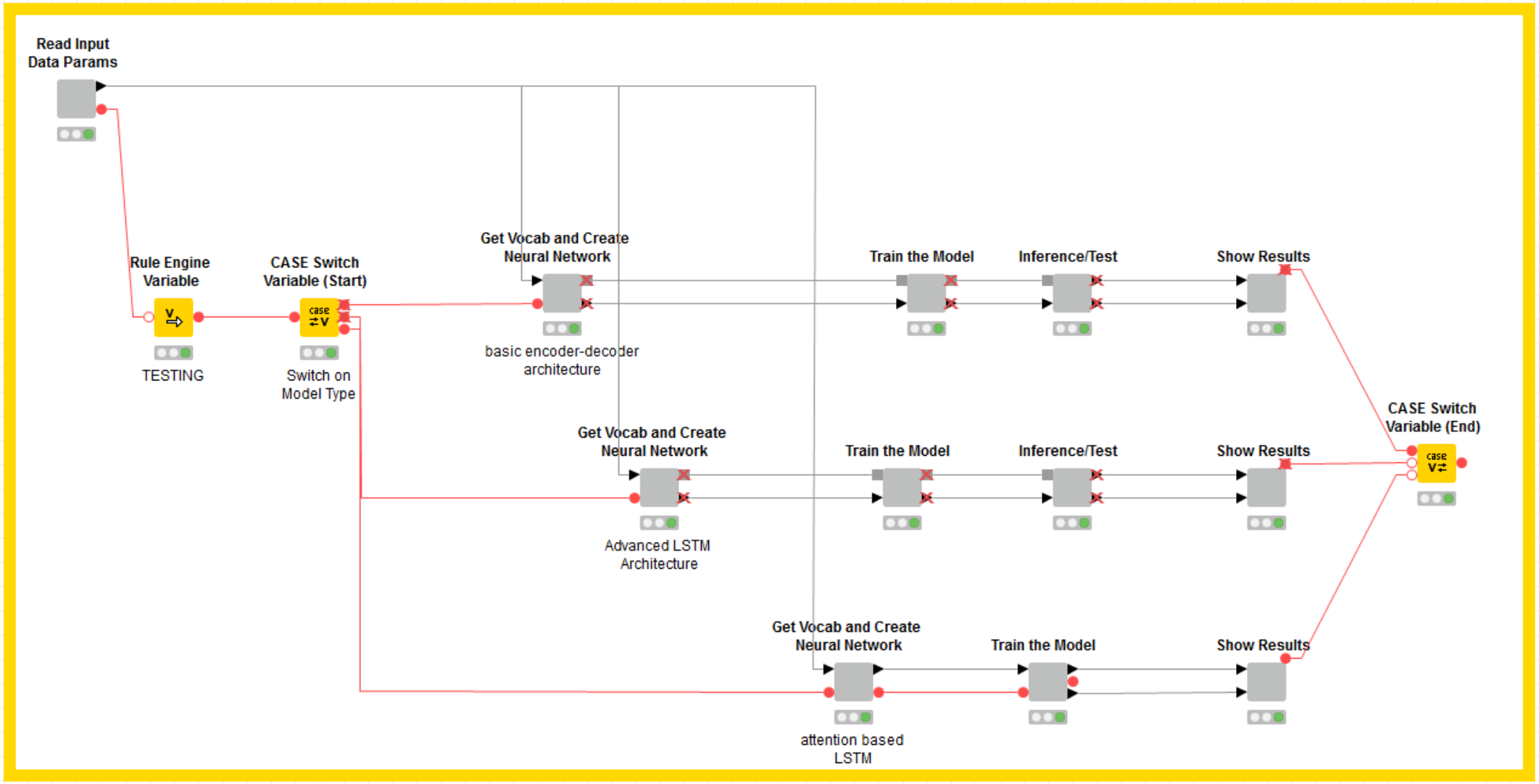
NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

Using KNIME for NMT

- Better integration, management, and documentation of workflows for internal and Cloud services
- Use KNIME Server to provide an easy user interface for parameter tuning
- Enable subject matter and language experts to review and edit NMT produced translations

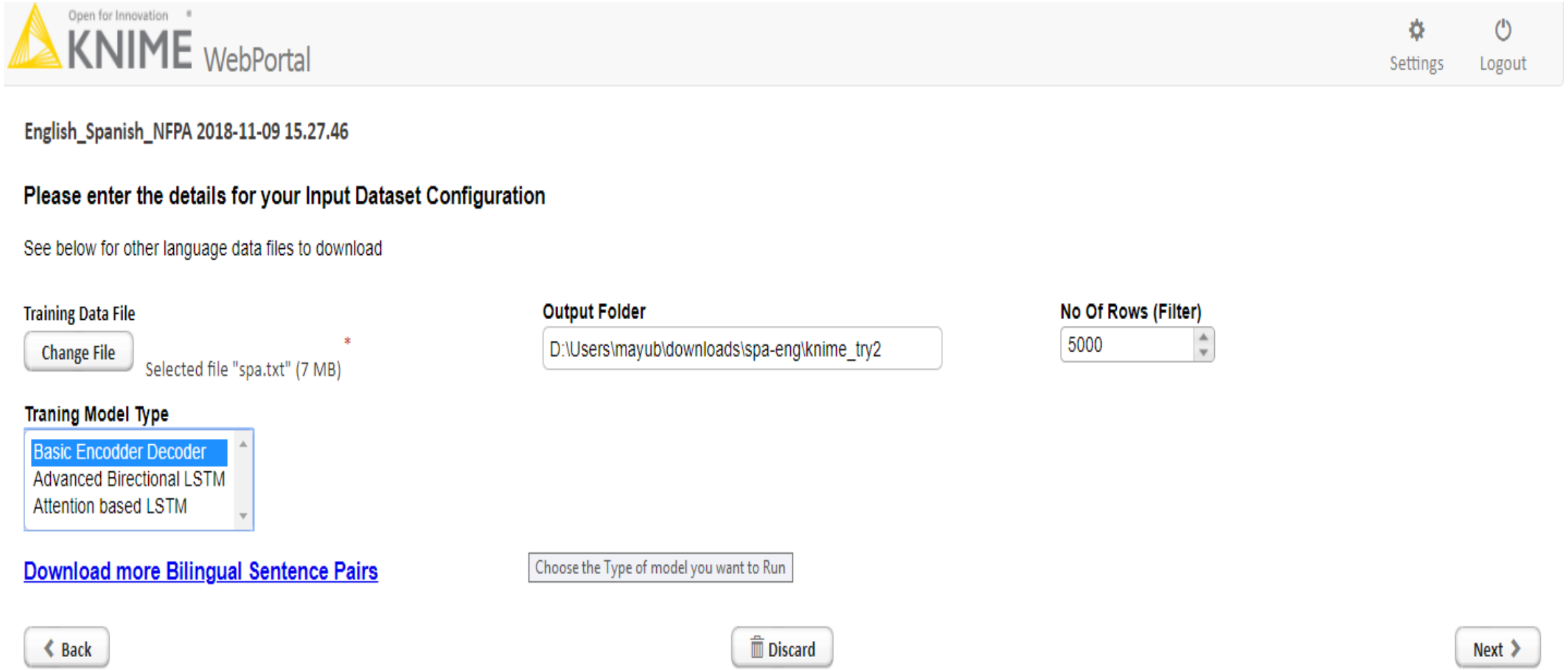
 © National Fire Protection Association. All rights reserved.

Keras-TensorFlow KNIME Pipeline



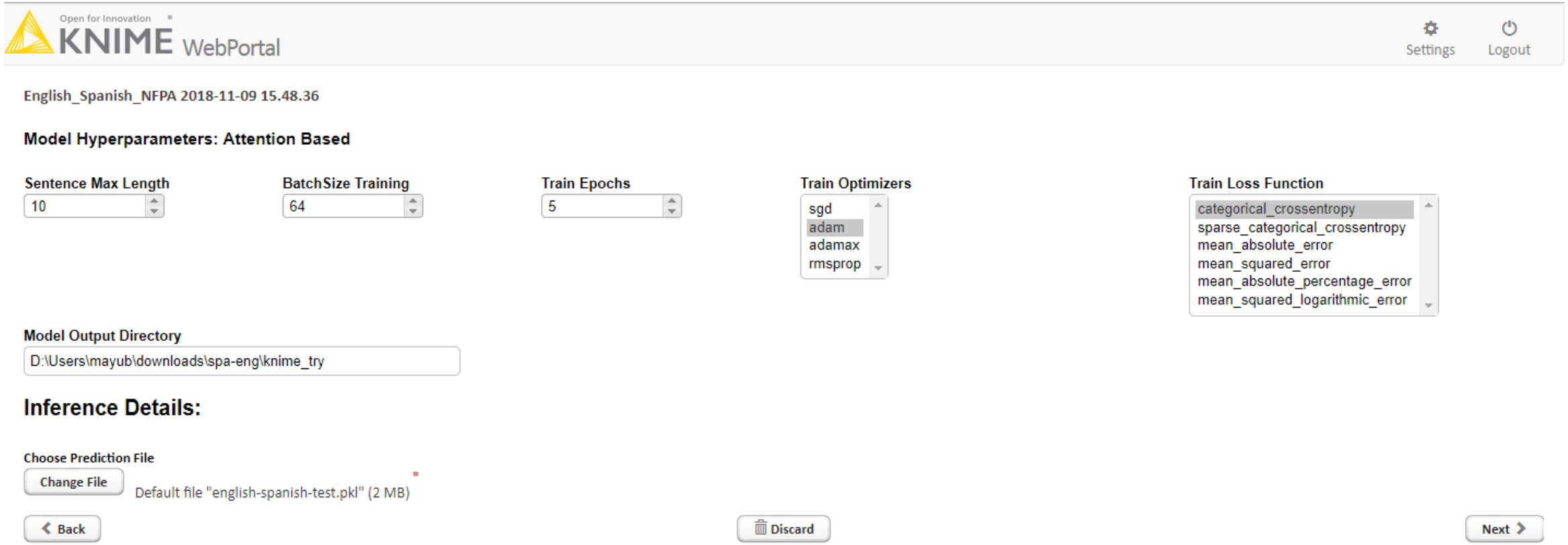
NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

Keras Pipeline (continued…)



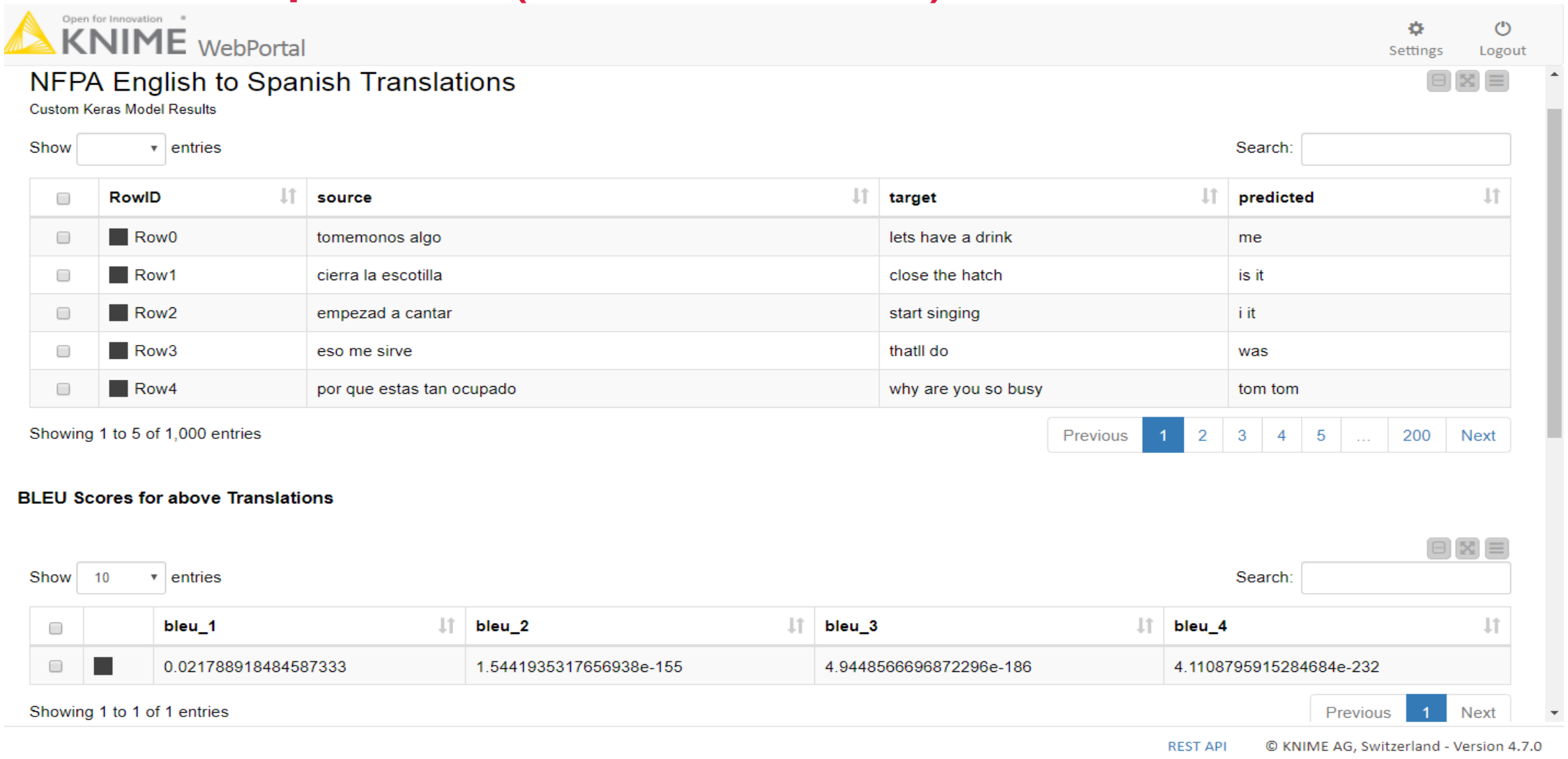
NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

Keras Pipeline (continued…)



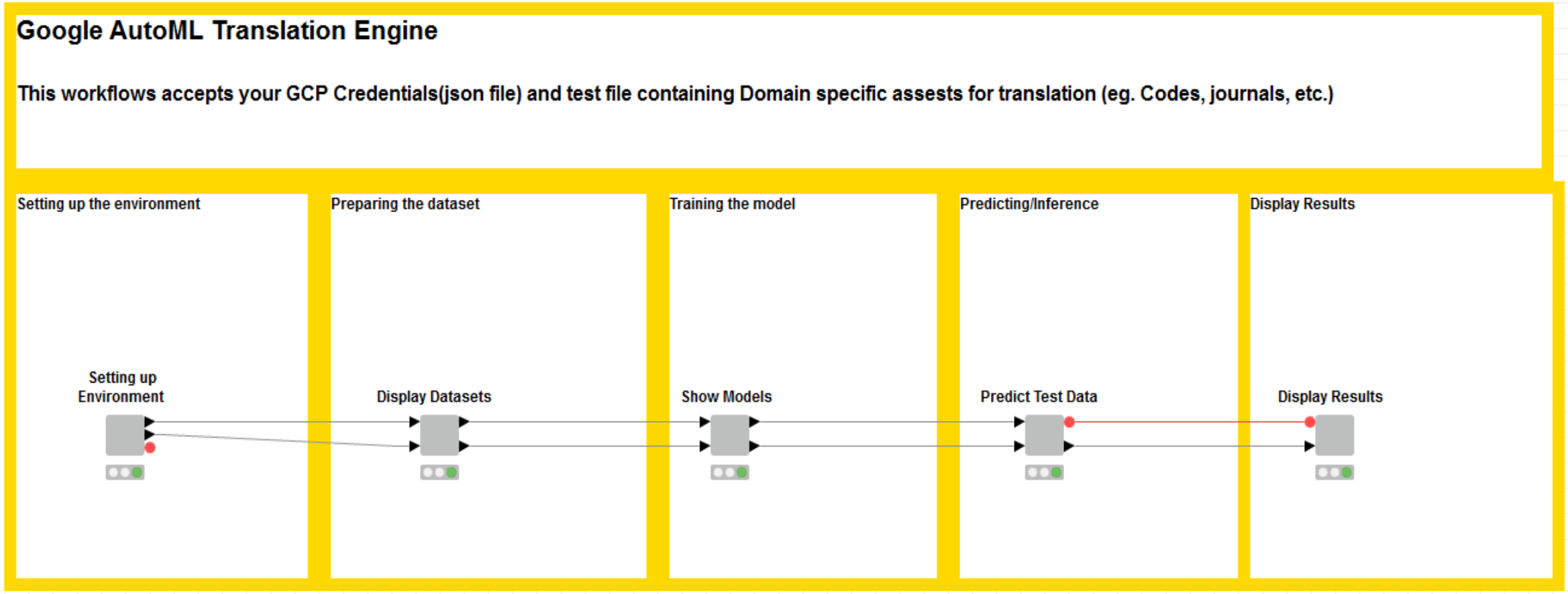
NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

Keras Pipeline (continued…)



NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

AutoML KNIME Test Pipeline

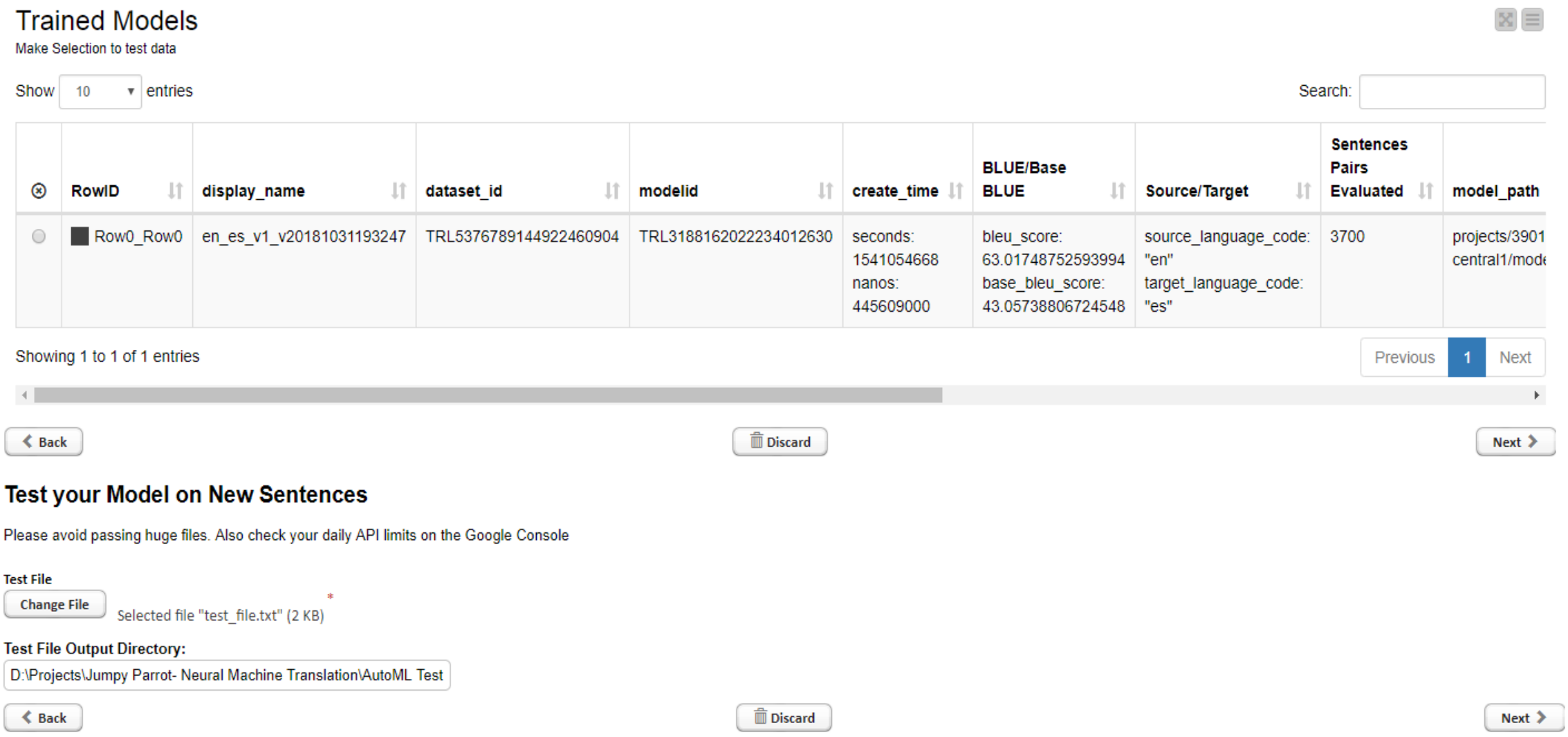


NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

AutoML Pipeline (continued…)

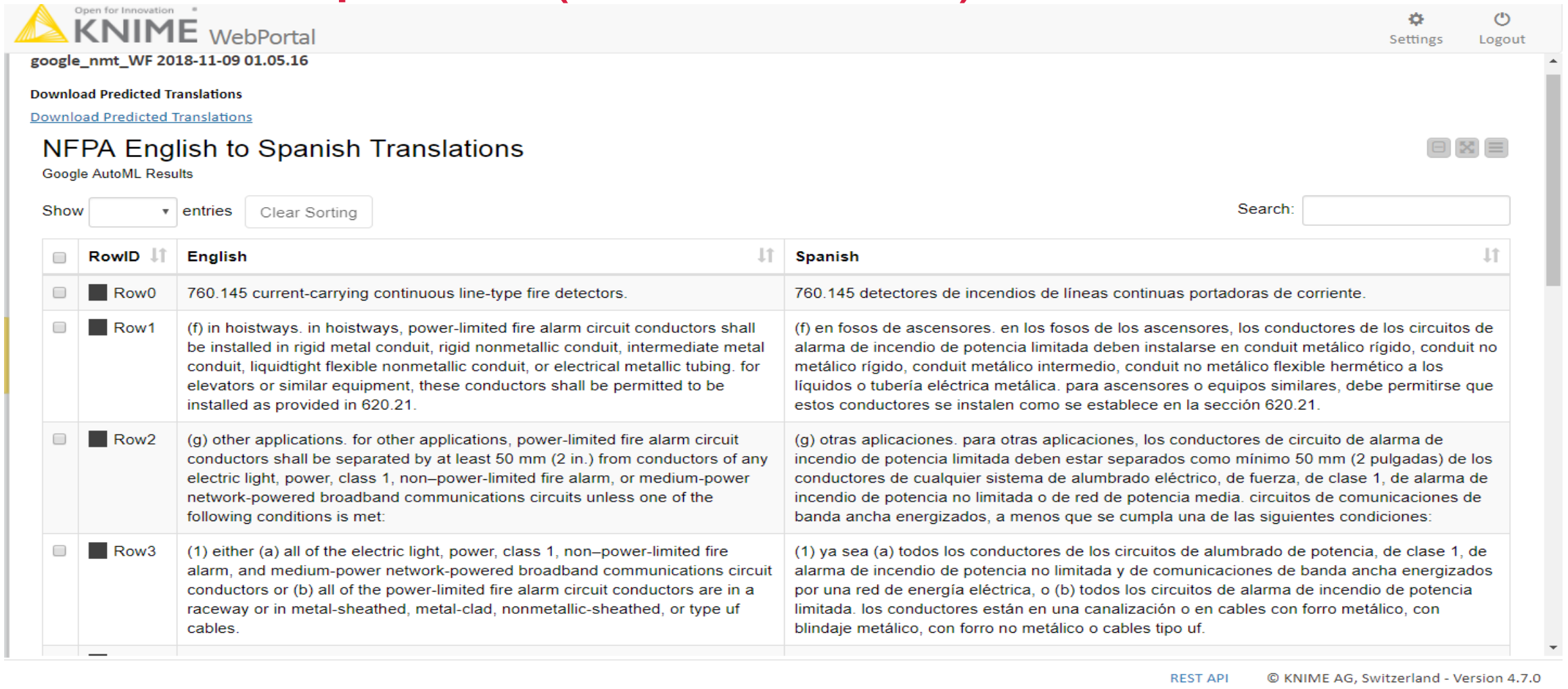
 © National Fire Protection Association. All rights reserved.

AutoML Pipeline (continued…)



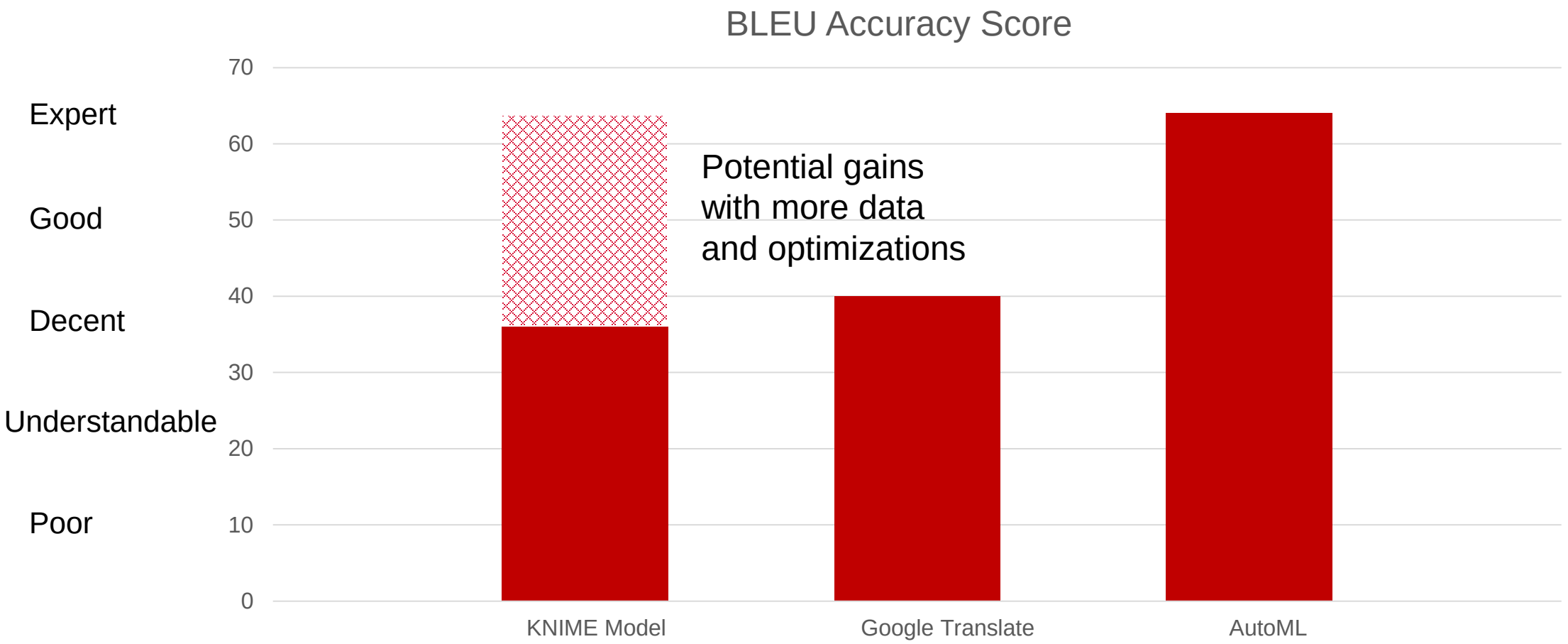
NFPA.ORG | NFPA Confidential. For Internal Use Only. | © National Fire Protection Association. All rights reserved.

AutoML Pipeline (continued…)



 © National Fire Protection Association. All rights reserved.

Results



 © National Fire Protection Association. All rights reserved.

Results –

	BLEU (w/o training on NFPA Data)	BLEU (with training on NFPA Data)	Performance Gain (BLEU points)
Auto ML	43.1	63	~20
KNIME models	12.5	36.0	~24

 © National Fire Protection Association. All rights reserved.

Results –

for further information for fire alarm systems, see article 760.

AutoML Custom:

para mayor información sobre los sistemas de alarma de incendio ver el artículo 760.

Official NFPA Translation

para mayor información sobre los sistemas de alarma de incendio, ver el artículo 760.

KNIME Model

para mayor información sobre los sistemas de alarma de incendio, véase el artículo 760.

 © National Fire Protection Association. All rights reserved.

Example – Technical Terminology

NFPA 1, section 12.7.4.2.1

Fire-rated glazing assemblies marked as complying with hose stream requirements (H) shall be permitted in applications that do not require compliance with hose stream requirements.

AutoML:

Se deben permitir ensambles de vidrios resistentes al fuego marcados como que cumplan con los requisitos de la corriente de la manguera (H) en aplicaciones que no requieran cumplir con los requisitos de la corriente de la manguera.

Official NFPA Translation

Deben permitirse conjuntos de montaje de vidrios certificados como resistentes al fuego señalizados, lo que indica que cumplen con los requisitos para chorros de manguera (H), en aplicaciones en las que no se requiera cumplir con los requisitos para chorros de manguera.

KNIME Model

se debe permitir que los productos de vidrio, que cumplan con los requisitos de la corriente de flujo, se utilicen en aplicaciones que no requieran el cumplimiento de los requisitos de la corriente de flujo

 © National Fire Protection Association. All rights reserved.


takk
shukran
xièxie
dank-u
kop-khun
danke
efharisto
spasiba
kiitos
kamsahamnida
grazie
merci
ThankYou
salamat
mahalo
arigato
terima-kasih
hvala
gracias
obrigado
tak