

Leveraging Neural Networks and Python to Forecast Train Delays in the Swiss Railway Network

Anpassung mit RL gem. Konzept und BLM folgt."

311,"K,mmer BZ Mitte(25 02 21) im Einsatz"

312,"71614 (ex 1611) wird zur Zeit in MST ger%eumt."

313,"Gem. ZP Zug ger%eumt... Zug fahrbereit"

314,"Unfallzug befindet nach der Haltestelle BSDH Richtung BS auf h'he Tramdepot."

315,"KUBE en gare de BI"

316,"MOCO BI mit Erstinformation aktiviert."

317,"Merci de dire aux AT de faire une annonce avant BI."

PAX BS via OL."

318,"Kdt. BW vor Ort. Einweisung Blaulichorganisation durch BW"

319,"So lange keine Konzeptphase werden Kunden mit Ziel Laufental mit Tram 10 nach Dornach gelenk
Tram 11"

320,"Evakuierung mit LRZ."

ca. 100 Pax im Zug"

321,"BLM BI: bitte anpassen auf ""Es ist mit Versp%etungen und Zugausf%allen zu rechnen"". Merci"

322,"Evakuierung nach DSDH. Reisende ben,tzen anschliessend Tram nach BS"

323,"In Absprache mit LKR BZ Mitte / LKR OCP:

Entscheid f,r Umsetzung Konzept um 09h30.

Konzeptidee:

- ICN 16xx wenden in AE / ab BS Ausfall

- S3 173xx wenden in AE und BS"

324,"Evakuierung nach mst. Reisende ben,tzen anschliessend Tram nach BS"

325,"SIWf vor Ort eingetroffen"

326,"Lf,r <berf,hrung Unfallkomp folgt ab BS Depot K1 (Dispo Lf) ist in ca. 10.Min vor Ort"

327,"@Newsroom

Sollen Reisende ins Laufental ab BS nach Dornach oder nach M,nchenstein gelenkt werden?"

328,"verunfallte Person leicht verletzt"

329,"In Absprache mit BZ Mitte verkehrt 1613 durch bis BS, ev. leichte Versp%etung."

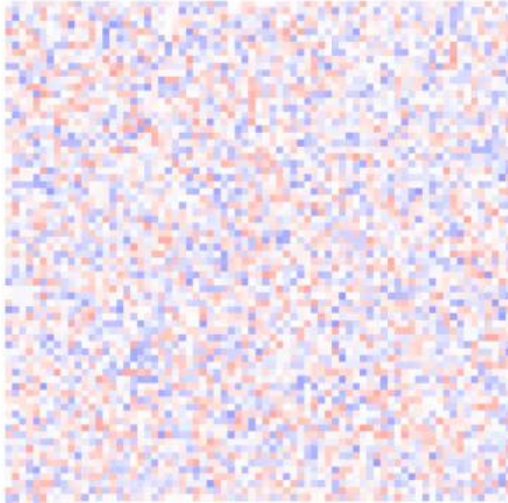
330,"Evakuierung l%auft"

331,"KEIN Konzept - Einzelmassnahmen durch BZ Mitte:

2016-07-31 09:00:33



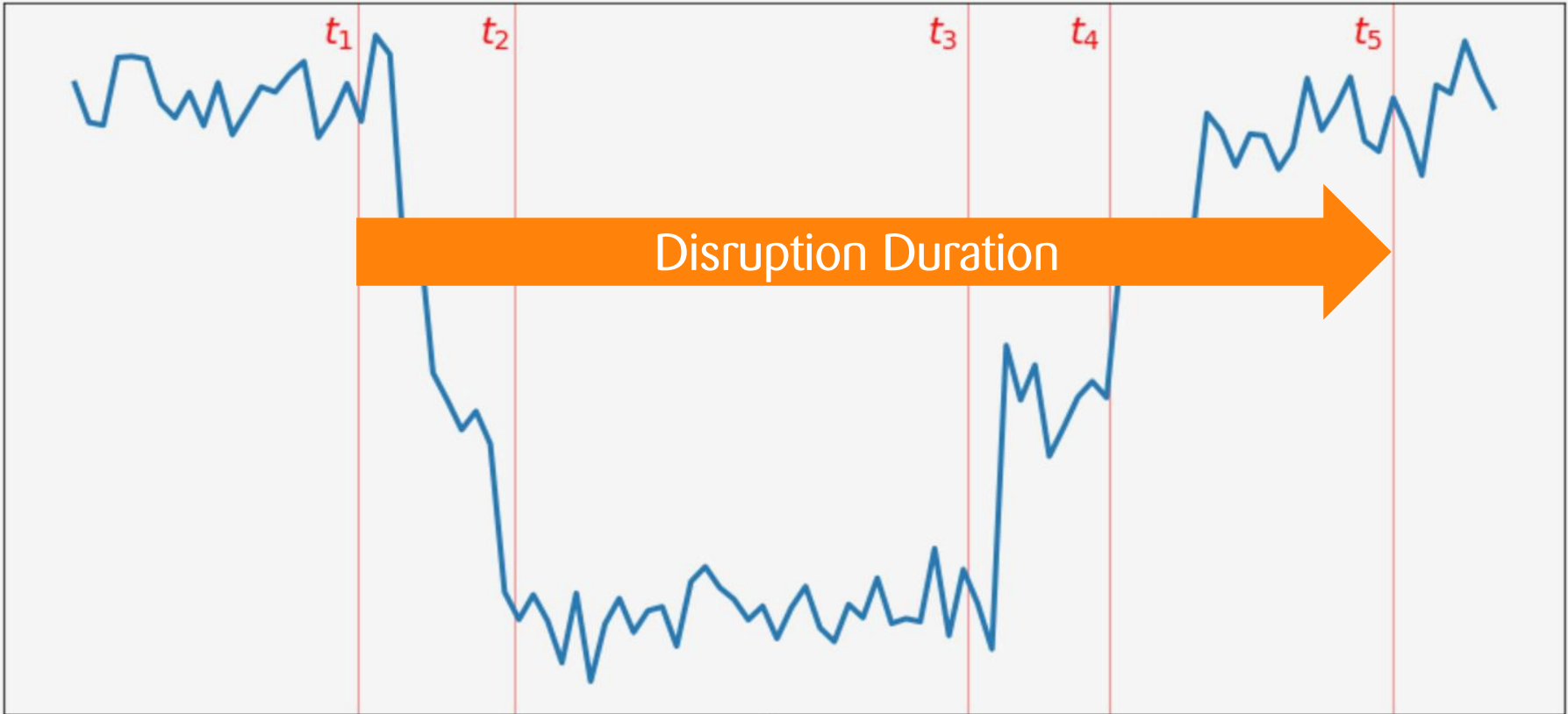
Accuracy



Overview

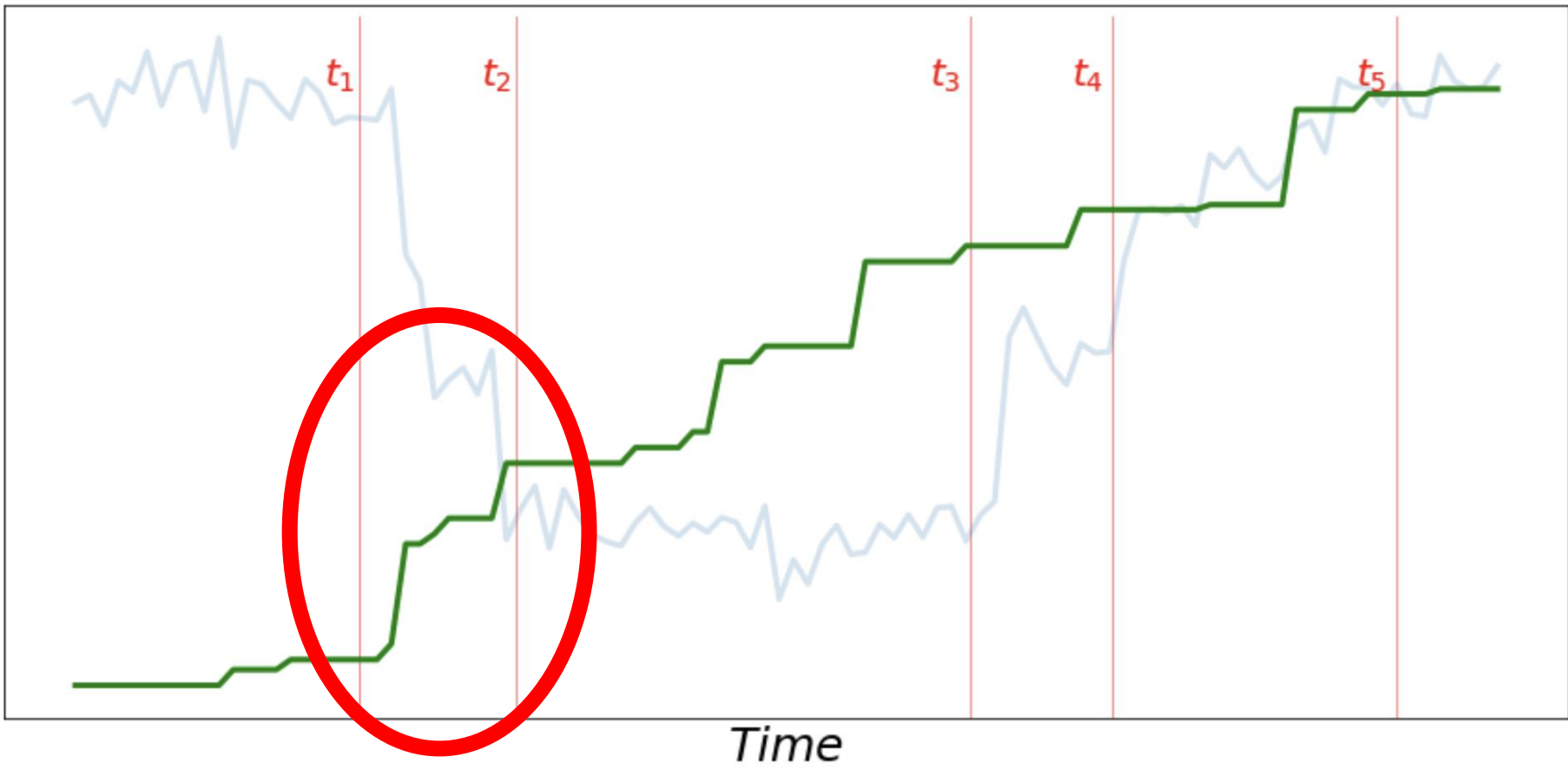
1. What is a Disruption?
2. Machine Learning with Text Data
 - Text Features
 - Sequence Models
3. The Disruption Prediction Model
4. Outlook

SBB Productivity



Time

Disruption Information



u150256 (Voyeur / Standard)

Verantwortungsbereich

BZ Mitte - Voyeur

Zusätzliche Bereiche

Übersicht

Fallübersicht

				Ort	Zug	Ereignis	ErZu-Nummer	Erstellt	Geändert
1	1			FR	43564	Personaldispos... Folgefall PU	✓	10:25	10:52
	2			ZUE	8	Fahrzeugstörung - Lokstörung		10:23	10:37
2	1			SO		Person / Tier i...n in Gleisnähe	✓ 2135127	10:15	10:57
1	1			W-ST5	40788	Geänderte Fah...eu via BUE-KB	✓	10:10	11:25
1	1			RBLA	62228	Zugbereitstell...agen 33823547	✓ 2135128	10:08	10:38
	1			LZ-BRUE		ungültiger Fall	✓	10:00	10:04
1	1			LN	97531	Fahrzeugstörung	✓	09:56	10:29
3	3			ROSS	15533	Zug fährt nicht		09:47	10:07
15	24		2	SIS-GKD	99411	Fahrzeugstörung	2135102	09:46	10:33
2	1			BN	810	Polizeiinsatz - TPO +2	✓	09:44	10:43
	5			GUE		Stellwerkstörung - W47		09:43	10:42
1	2			SIS	99411	Fahrzeugstörung - Reset	✓	09:41	09:43
1	1			RBLA	62124	Personaldispo...rsp. ab Depot	✓ 2135123	09:34	10:36
3	4			LTS	60565	Fahrzeugstörung - dgt loc	✓ 2135109	09:34	11:11
	4			WR-UT	16632	Unregelmässig...lag verspürt	✓ 2135104	09:34	10:27
1	1			CHV	50857	Fahrzeugstöru...rangement loc		09:33	09:34
1	2			LTS	60565	Fahrzeugstörung - dgt loc	✓	09:29	09:31

Anfragen

Beschreibung	Reisende	Bem
31462 in BS		
• 9210	20	
2345 in LZ		
• 21592	2	Zield
809 in BN		
• 709	80	ZUE

15 Massnahmen 24 Meldungen Verständigung 2 ECL Betrieb 2 Reisende

Meldung	Ort, Bereich	Zug	Ersteller	Erstellt	SMS
• 99411 ist am Dateneinlesen von Bombardier. Dauer mindestens eine Viertelstunde.			BZ M...rs	10:33	
• LF kann momentan mit 30km/h fahren, jedoch macht er noch weitere Tests in SIS Gl. 2. Probleme der Störung sind unbekannt			BZ ...rs	10:16	
• Alarmierung102-CLRZ-Alarm INT: SBS-OL-C-LRZ: Alarm 99411 aufgehoben /fr			ALS... CH	10:16	
• Globale Ereignisinformation			OCP... 2	10:15	

Die Störung zwischen Sissach und Gelterkinden auf der Linie Basel

Calculate travel time with Google Maps

```
import googlemaps
converter = GPSConverter()
converter.LV03toWGS84(dx['y'], dx['x'], dx['h'])
gmaps = googlemaps.Client()
gmaps.directions(adress, lat_lon, start_time)
```

Forecast Disruptions with Machine Learning

Disruption

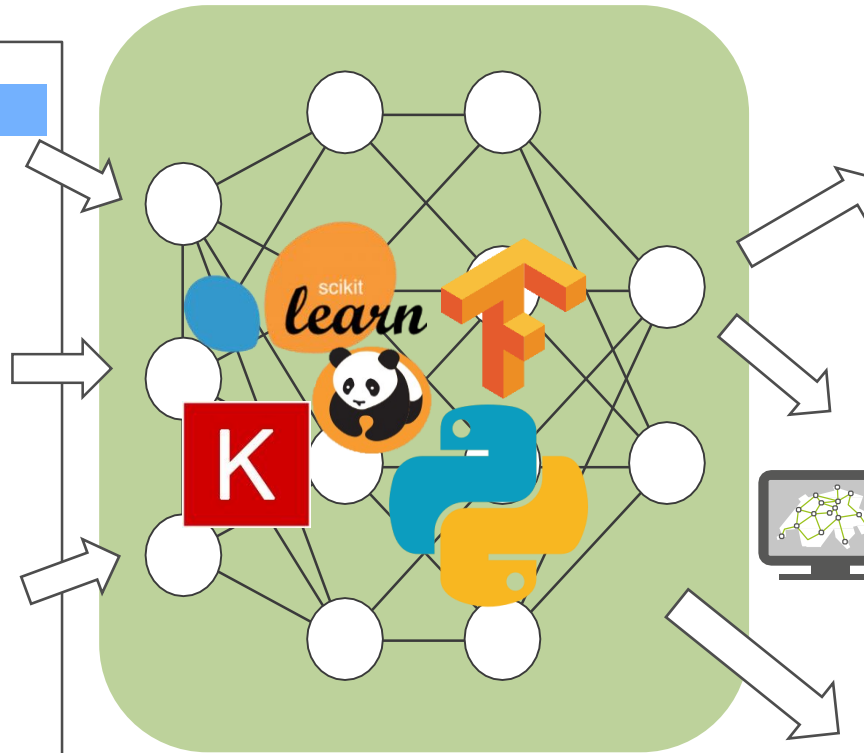
16:11 Stellwerkstörung «Weichenstörung» BRBA

Time **Message**

- 16:11 «Weichenüberwachung fehlt **Weiche** 122a»
- 16:12 «Keine Fahrten mehr möglich von Basel BAD nach BRBA (ausser A15/A16)»
- 16:13 «**TEB** verständigt»
- 16:14 «Fahrten von **Basel GB** nach **Basel RB /A/ Muttenz** nur mit **Hilfssignal** möglich (**Schutzweiche**)»
- 16:20 «Erste Technische **Prognose** kann um **16:45** erwartet werden»
- 16:36 «**Pikett** vor Ort»
- 17:18 «Fachdienst meldet erweiterte Störungssuche»
- 18:24 «Störung von selber verschwunden vermutet wird ein wackelkontakt **Fachdienst** wartet noch kurz ab»
- 18:29 «**Pikett** meldet sich ab Ende der Störung»

18:32 End of Disruption

Machine Learning



Disruption Duration



Train Scheduling



Customer Information



Supervised Text Classification



Training data

Test data



'Letter'



'Blog'



'News'

X_{train} Y_{train}

Label

```
classifier.fit(X_train, y_train)  
keras_model.fit(X_train, y_train)
```



Learning Method

Classifier

Prediction

$f: X \rightarrow Y$

Prediction function



?



?



?

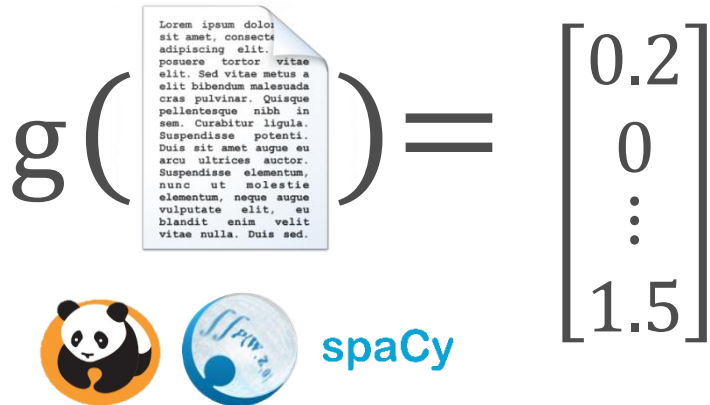
X_{test}

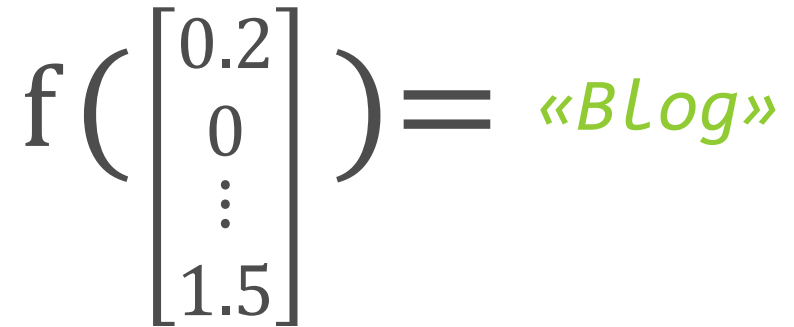
```
classifier.predict(X_test)
```

Machine Learning with Text

1. Create Feature Vector from Text

2. Train Classifier on Feature Vectors


$$g(\text{document}) = \begin{bmatrix} 0.2 \\ 0 \\ \vdots \\ 1.5 \end{bmatrix}$$


$$f\left(\begin{bmatrix} 0.2 \\ 0 \\ \vdots \\ 1.5 \end{bmatrix}\right) = \text{«Blog»}$$

```
X = numpy.zeros((n_documents, n_features))  
for i, doc in enumerate(documents):  
    X[i,:] = g(doc)
```

```
model.fit(X_train, y_train)  
y_pred = model.predict(X_test)
```



Text Features: TF-IDF, Bag of Words

Doc1: «Python is the best!»
Doc2: «Python is great for ML»
Doc 3: «Python is great»

g: TF-IDF

- How frequent is a word in a document?
- How frequent is a word overall?

g: Bag of Words

Term	Doc1	Doc2	Doc3
Python	1	1	1
great	0	1	1
the	1	0	0
best	1	0	0
is	1	1	1
for	0	1	0
ML	0	1	0

1. No word similarity
2. No sentence structure

```
bow = CountVectorizer()  
X_train = bow.fit_transform(train_docs)  
  
tfidf = TfidfVectorizer()  
X_train = tfidf.fit_transform(train_docs)
```



Dense Word Embeddings

Word Embedding

$$e(\text{«great»}) = \begin{bmatrix} 0.1 \\ 2.4 \\ \vdots \\ 1.8 \end{bmatrix}$$

Dense Word Embeddings

Word Embedding

“One-hot” vector

$$e(\text{«great»}) =$$

$$1. \text{onehot}(\text{«great»}) = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

Dictionary
Python
great
the
best
is
for
ML

$$2. \text{embedding} \left(\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 0.1 \\ 2.4 \\ \vdots \\ 1.2 \end{bmatrix}$$

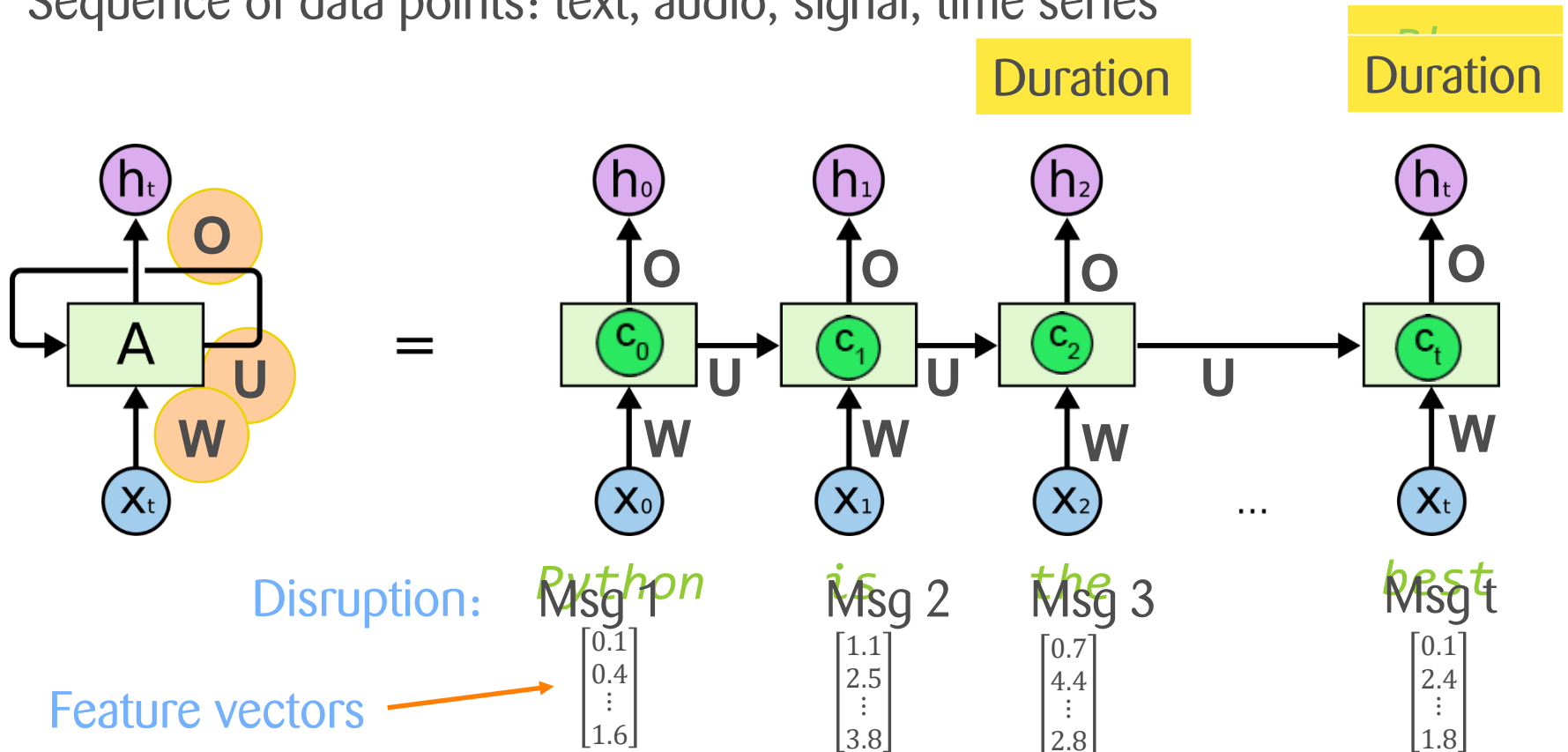
- word2vec: *great* ~ *best*
- Joint supervised

No sentence structure

Dense feature vector

Recurrent Neural Network

Sequence of data points: text, audio, signal, time series

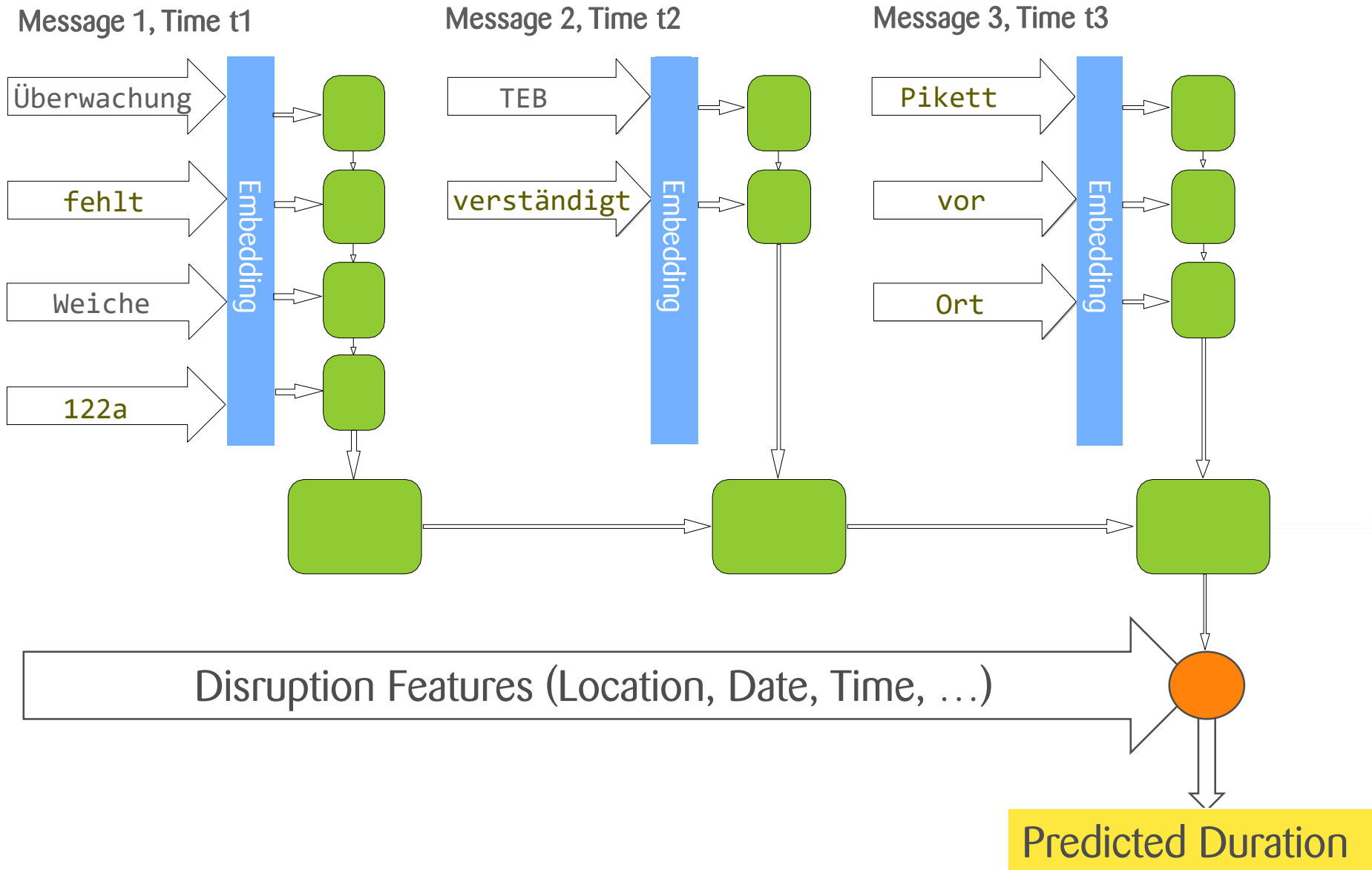


```
input = Input((1_sequence, n_features))
output = LSTM(n_hidden)(input)
model = Model(input, output)
```



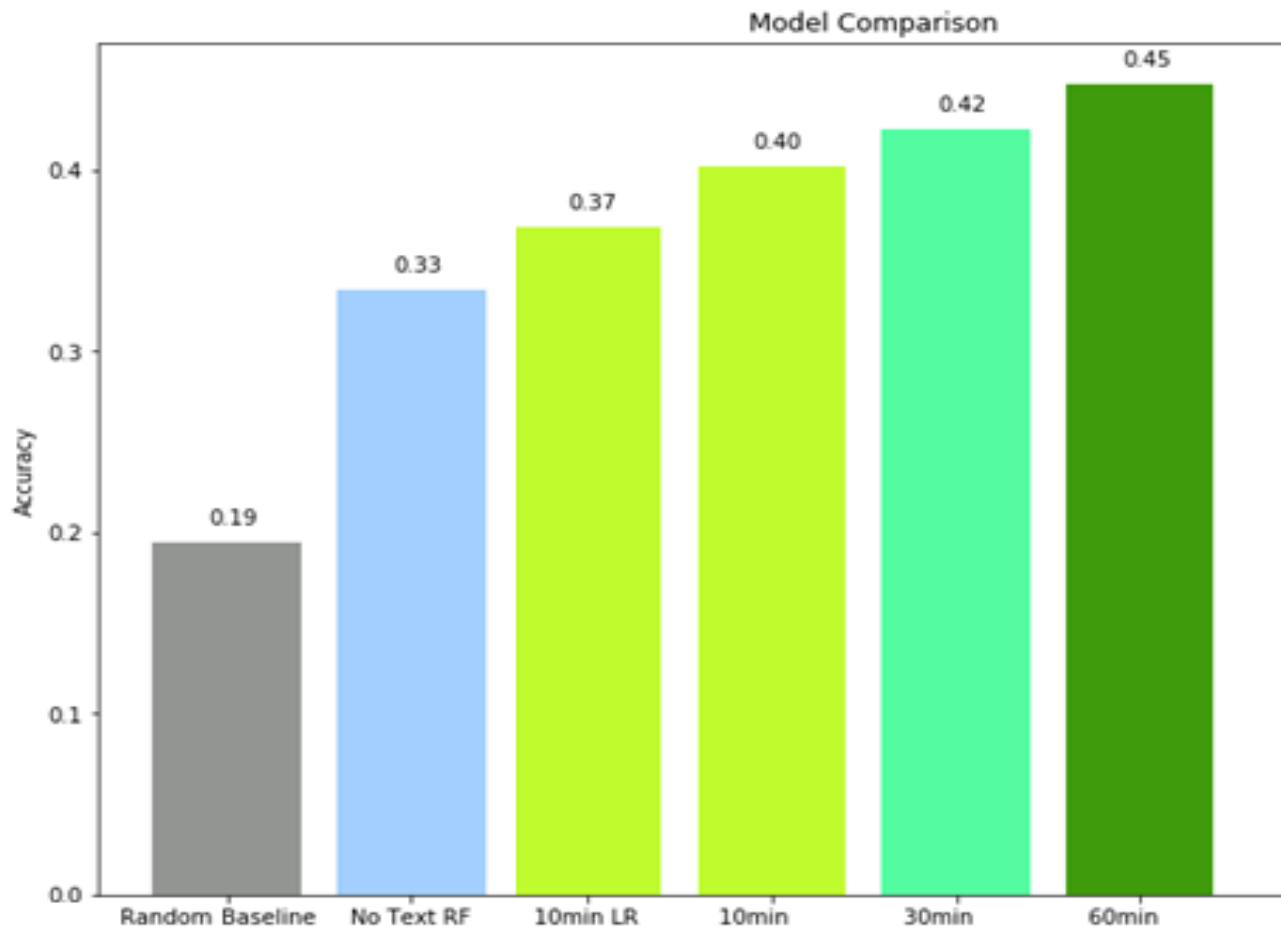
* Christopher Olah, colah.github.io

Putting it together: LSTM over Words and Messages



Results so far

- One type of disruptions in German-speaking Switzerland for one year
- TF-IDF per Message, LSTM over messages



6 Classes

0 - 15
15 - 30
30 - 60
60 - 90
90 - 120
>120

minutes

Product Vision

Questions?