

# Connecting the Dots: Matching Existing Solutions to New Production Defects



Team: Chloe Lee  
Daniel Cao  
Advisor: Retsef Levi  
Sponsors: Maka Karalashvili  
Virendra Pathak

## Problem

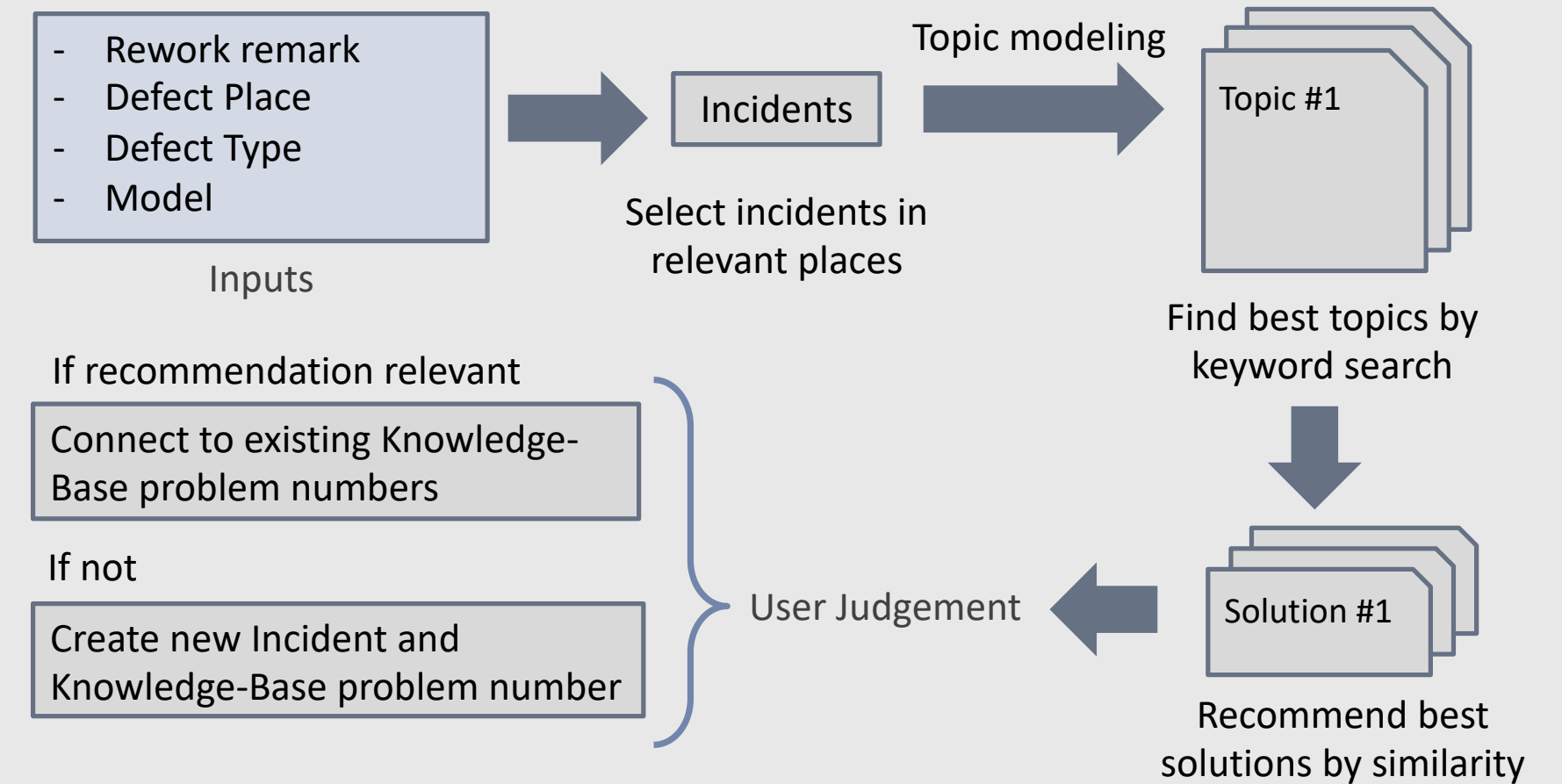
Defects occur frequently on the production line and quality engineers would have to address them on a daily basis. Valuable time are spent to fix these defects so similar defects are aggregated into the Knowledge-Base along with a common solution. However, this is a manual process that requires up to 24 hours per month of labor, and it is very prone to human error.

- How can BMW Group increase efficiency in the Problem Management Process?
- What's the most relevant solutions to a new defect?
- Does the solution exist in the current Knowledge-Base?



## Solution

A recommendation tool that suggest existing solutions to new defects



## Problem Management Process

The PMP system is used to manage defects and is consisted of the following three steps:

### 1. Identify Defects

Production defects are reported plant-wise, along with a short remark describing the defect in German and/or English.

### 2. Analyze Incidents

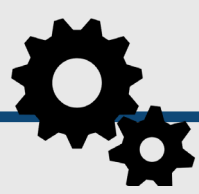
Critical defects are further analyzed and elaborated into incident descriptions.

### 3. Solve Problems

Similar incidents are assigned a unique Problem Number along with a detailed problem solution.

## Data and Methods

Each step contains a respective dataset and we conducted analytical methods both on and across the three datasets.

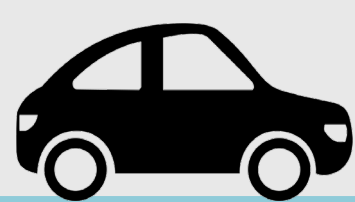


### Rework

- Over 4.3 million defect entries across different plants since 2019
- Structural features: defected vehicle parts and defect type
- Unstructured features: multilingual and succinct rework remarks
- Less than 0.05% entries are connected to a Problem Number

#### Sentence Transformers

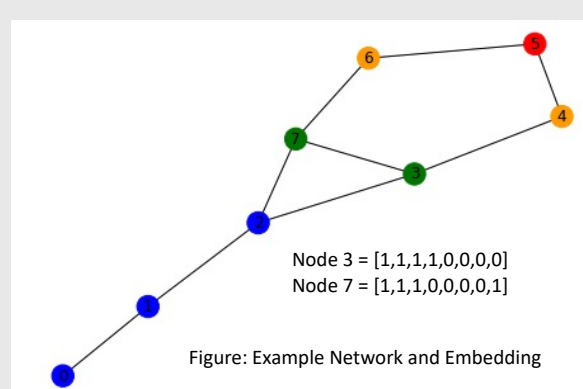
We utilized multilingual sentence transformers to generate embeddings of a new rework remark to understand the problem.



### Vehicle Parts Embeddings

**Challenge:** Overwhelming number of solutions to consider at once  
**Solution:** Segment incidents by their respective defected vehicle parts

- Generate parent node vector for each defect place
- Condense embedding vector through Principal Component Analysis (PCA)
- Cluster defected vehicle parts by their embeddings and match incidents to their respective defect part cluster



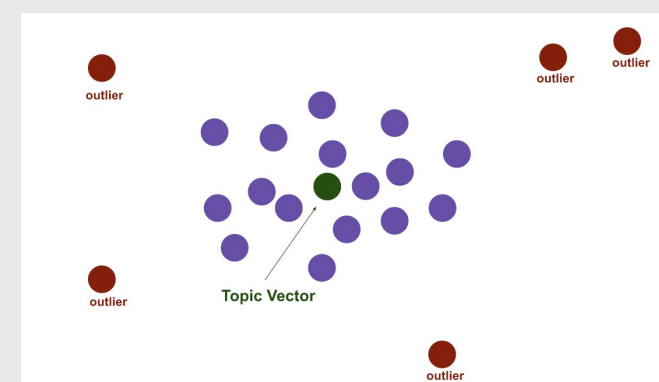
### Incident

- Around 1.1 million incidents
- Over 80% incidents are connected to a Problem Number
- Contains more detailed multilingual description about the defect

#### Topic Modeling: Top2Vec

We clustered incidents by their topics.

1. Obtain embeddings of incident description
2. Conduct dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP)
3. Perform clustering using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)



#### Short Text Clustering: Movie Group Process

Large topics are further clustered using the Gibbs sampling algorithm for a Dirichlet Mixture Model (GSDMM).



### Knowledge-Base

- ~891 thousands solutions (Problem Numbers)
- Detailed multilingual description of both defect and solution

#### Cosine Similarity

Problem Numbers are ranked based on cosine similarity of sentence embeddings of Rework remark and solution title to assess the relevancy between the defect and proposed solution.

#### Rouge Score

Rouge score measures overlap of n-gram between Rework remark and solution description.

#### Optional Features:

- Vehicle Model
- Type of Defect

### Keyword Extraction

**Challenge:** Incident descriptions contain many irrelevant information  
**Solution:** Identify most important information by extracting keywords of each topic cluster

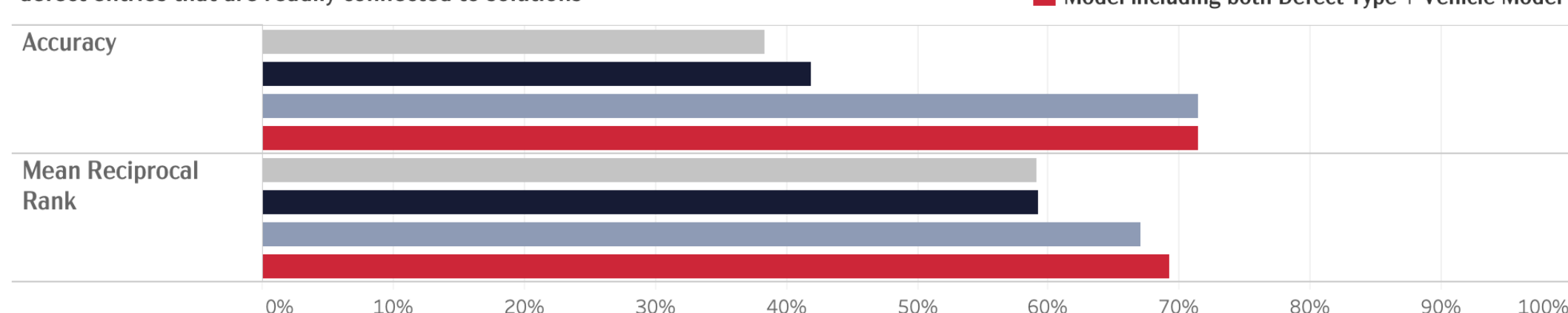
- Yet Another Keyword Extractor (YAKE!) Collection-Independent Automatic Keyword Extractor
- Regular Expression of Option Codes and Severity Index
- Extracting Noun-Verb pair through part-of-speech (POS) tags

**Objective:** recommend all Problem Numbers in topic clusters that contain relevant keywords to the respective defect

## Results and Impact

### Large-Scale Validation Result

The PMP Data Connection Pipeline tool is evaluated with defect entries that are readily connected to solutions



- The recommendation tool performs the best when it includes both Defect Type and Vehicle Model as additional inputs (accuracy: 71.43%, MRR: 69.23%)
- The defect solution recommendation tool relieves plant workers of cumbersome manual work and allows them to focus on roles that require more human expertise
- Machine learning also reduces human errors such as creating the same solution twice if they are unable to find the right solution among thousands of entries
- Improves efficiency by 900x while allowing flexibility in inputs to accommodate for user's domain knowledge

8K

projected defects  
matched per day

20min

work time reduced  
per operation

€ 500K

annual labor expense  
used more efficiently

9K+

manual labor hour  
optimized annually

30+

quality engineers  
directly assisted