Utilizing Word Embeddings and Gradient Boosting to identify, analyze, prevent and predict machine errors in the Computer Aided Manufacturing Industry

Aadit Kapoor

Index

- What are we trying to solve?
- About the Computer Aided Manufacturing Industry
- M.L Algorithms: Word Embeddings and Gradient Boosting
- Real world Impact

This is a Dental Milling Machine.

Dental mills are **designed to cut crowns, bridges, copings, frameworks, implant abutments and more from materials such as ceramics, zirconia, alloys, resins or wax**.

What are we trying to solve?

- For a dental milling machine, Can we achieve the following:
  - Prognosis: Identify/Analyze machine errors
  - Prevent: Prevent machine errors
  - Predict: Predict future error probability

About the Computer Aided Manufacturing Industry

- The Computer-Aided Manufacturing market size to grow from USD 2.3 billion in 2018 to USD 3.4 billion by 2023

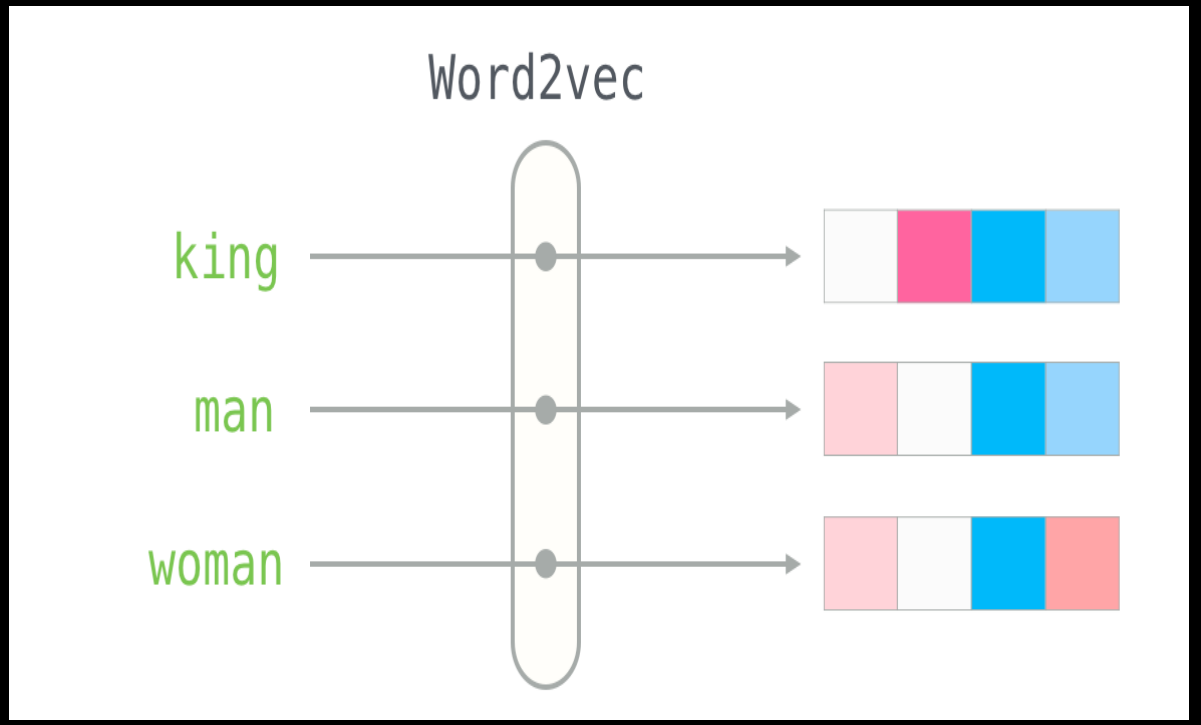
- Problem Statement: Can we use A.I/Data Science to supercharge the industry?
  - Better Customer Retention/User Experience
  - Faster Support Resolution
  - Design better machines (Eventually)

M.L Algorithms: Word Embeddings and Gradient Boosting

- Available Data: Sensor and Log Data
- Outcome: To predict future errors and build explainable models
- Dimensionality: 23 sensors. Typical case of class imbalance (0 denotes no error and 1 denotes error). Each sensor contributes to the functioning of the machine
- Data consolidation: Sensor Data and Log Data are separate data repositories.
- Things to consider:
  - Two variants of machines
    - Single Model for all machines
  - Separate Models
  - Compute Availability

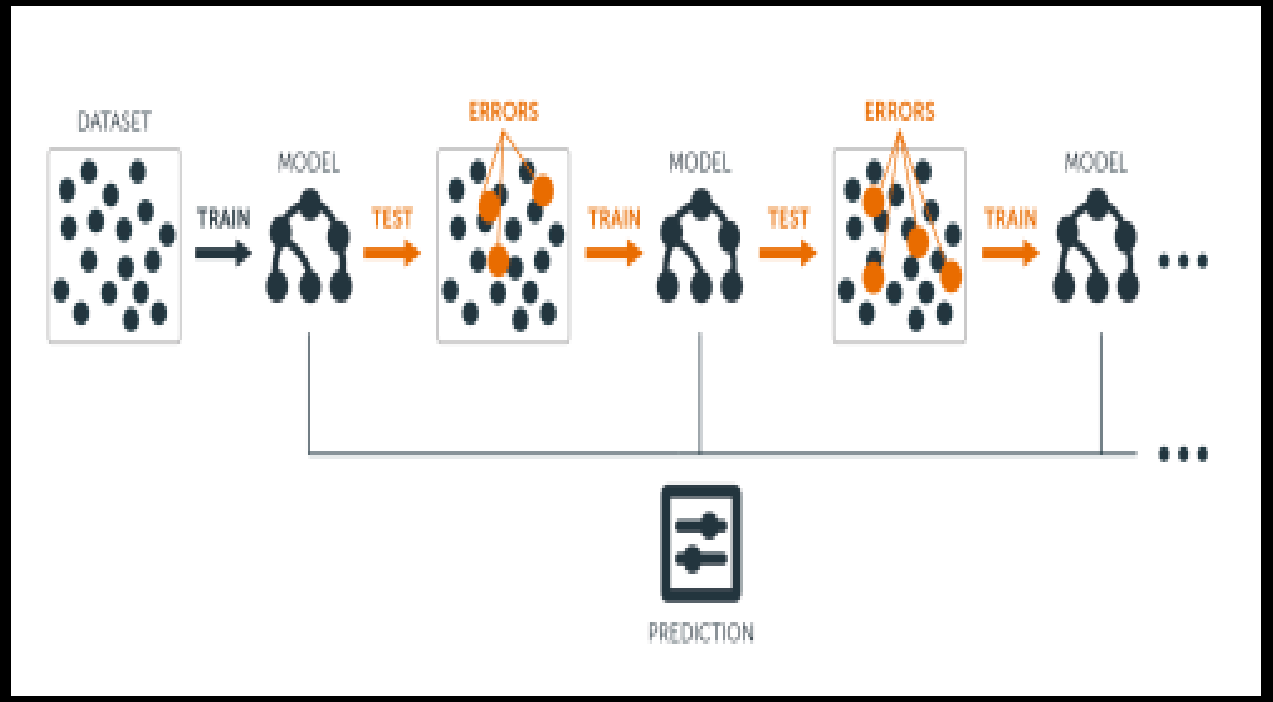
M.L Algorithms: Word Embeddings

- Word Embeddings
  - Convert Text log data into vectors to compute similar errors.
  - Advantages
    - Identify errors in unstructured data. Find frequency of errors
    - Map similar "days" for errors
    - Identify time of errors
    - And a lot more...



M.L Algorithms: Word Embeddings and Gradient Boosting

- Gradient Boosting
  - Utilize structured data and preprocessed text log data to perform gradient boosting to predict error prone sensors.
- Advantages
  - Combining word embeddings for sensors as features with sensor values
  - An end-to-end M.L training pipeline



Working of gradient boosted trees

Real world Impact

- What have we achieved?
  - An intuitive way to combine text data and structured data (tabular data) to predict error prone regions of a machine.
  - Utilizing word embeddings as a feature for gradient boosting increases performance by 10x
  - Show how A.I and data science can augment the Computer Aided Manufacturing industry by predicting errors beforehand

Thank you!