



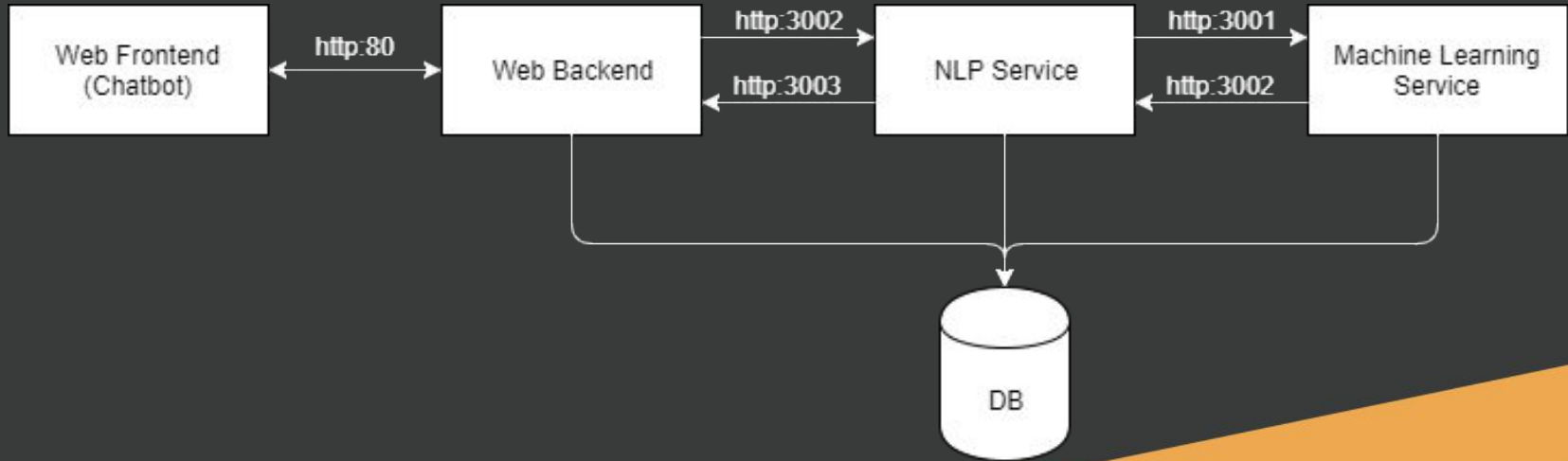
PROCE EUS

AI Legal Chatbot
(Release 2)

Presentation Overview

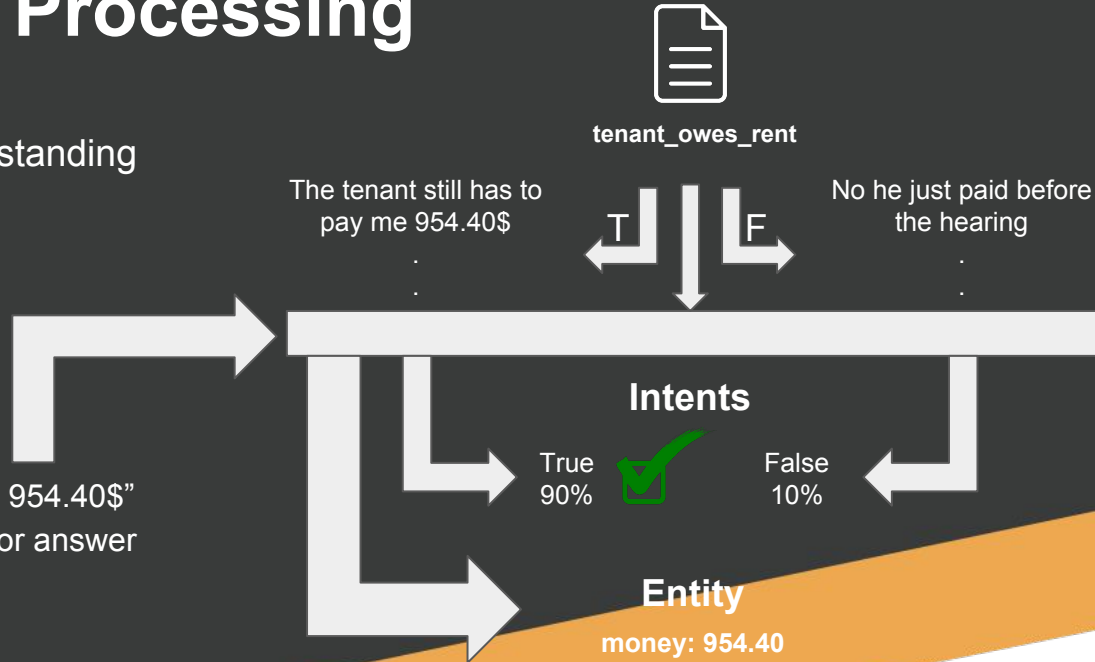
- Overview of project architecture
- Natural Language Processing
 - Custom Rasa Pipeline: `nlp_spacy`, `tokenizer_spacy`, `intent_entity_featurizer_regex`, `intent_Featurizer_spacy`, `ner_crf`, `ner_synonyms`, `intent_classifier_sklearn`, `ner_duckling`
 - RASA AI 0.11.0 - Using SpaCy 2.0 (on dev branch as of 1.31.2018)
- Machine Learning
 - Data Pipeline
 - Classifiers (SVM)
 - Regressors (SVM + Neural Network)
- Demonstration

Overall Architecture



Natural Language Processing

- Framework
 - RASA Natural Language Understanding (soon 0.11.0)
- Basic Flow
 - Bot asks question:
 - “Does your tenant owe you money for late rent?”
 - User answers:
 - “The tenant still has to pay me 954.40\$”
 - NLU System generates confidence for answer
 - Fact: `tenant_owes_rent`
 - True - confidence 90%
 - False - confidence 10%
 - Entity: money (using `ner_duckling 1.7.2`)
 - money: 954.40



Natural Language Processing

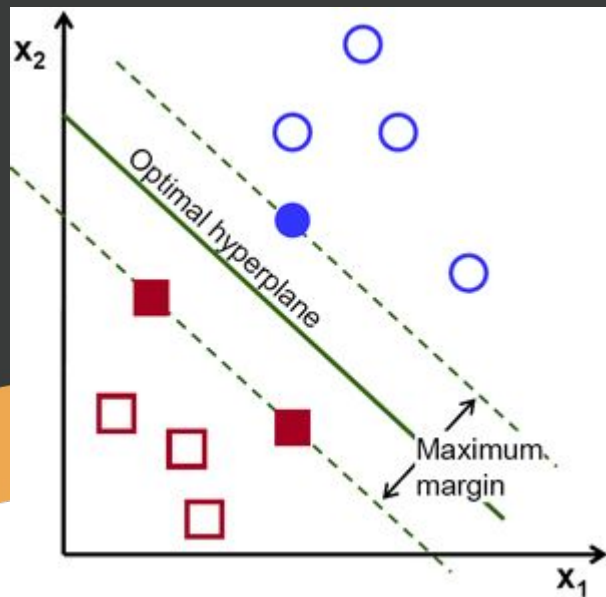
- For 25 frequently-asked user demands:
 - Immediate classification and response by NLP
- For 2 more elaborate user demands (lease termination, nonpayment):
 - Train intent classifier models for every fact relating to each topic
 - Identify user intent, possibly extract entity, save fact, ask ML for next question / prediction
- Simple tool developed to parse example sentences and provide input to RASA: greatly simplifies training of ~40 classifiers

Machine Learning

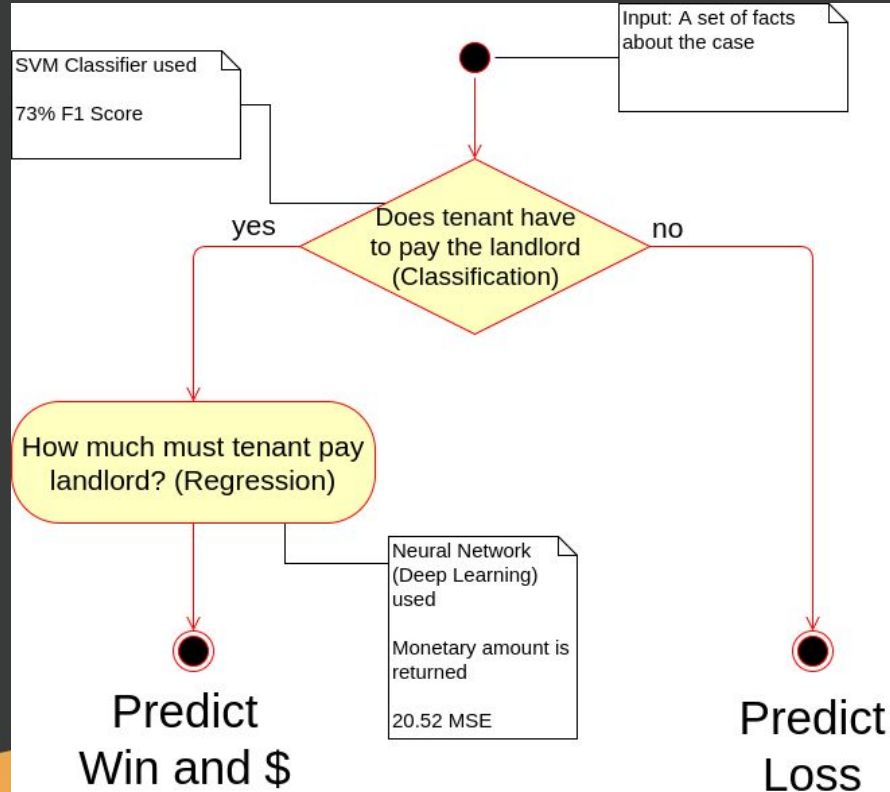
- Current Statistical ML Predictions
 - Classification
 - Tenant Expulsion
 - Lease Termination
 - Regression
 - Monetary amount that the tenant must pay the landlord (Classification & Regression)
 - Similarity
 - Finding similar precedents to your case

Prediction: Lease Resiliation

- Support Vector Machine
 - Relies on finding the best hyperplane, in this case line that “best” separates the different classes
- 96% F1 Score



Prediction: Amount tenant owes landlord



Prediction of amount tenant owes landlord

- Inputs: Facts pertaining to the case
- Output: Total monetary amount that is owed to the landlord (punitive fees, judicial fees, compensatory damages, rent)
- Areas of Improvement:
 - Likely that the input fact set is incomplete. Manual review of clusters and case proceedings can improve our fact set.

Precedent Similarity

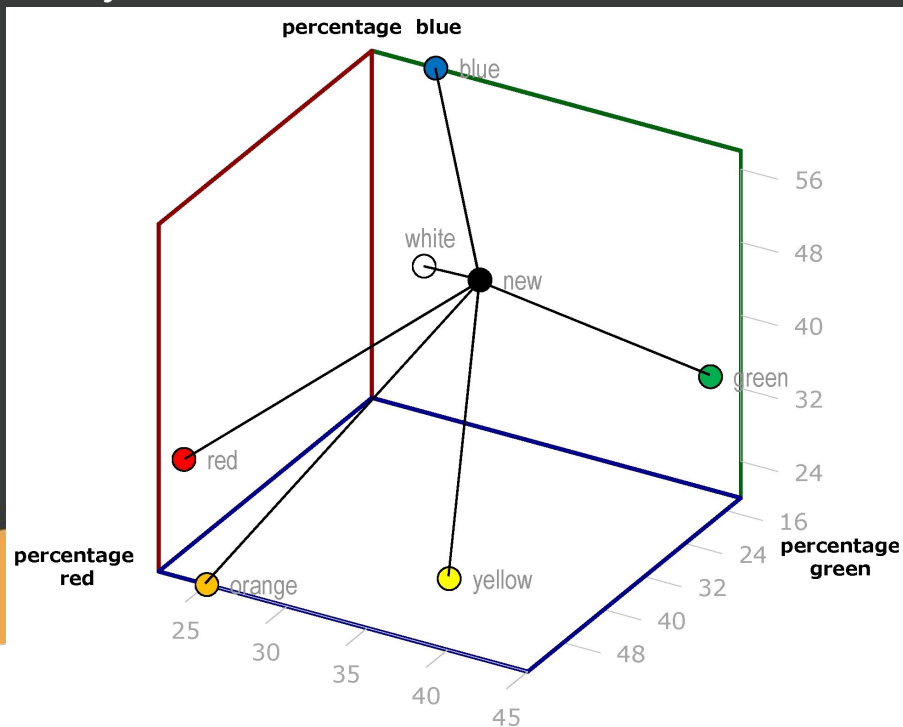
- Determine which precedents are similar to your case
- Experimenting with a few similarity

Metrics:

- Cosine Similarity
- Nearest Neighbour with mahalanobis distance

Limitations:

- No access to evaluation metrics



Prediction Overview

Support Vector Machine

- 1 estimator to predict multiple outputs
- value : 1 indicates the presence of an outcome
- Value: 0 indicates the absence of an outcome (Null / None)
- E.i.: 'lease is terminated' AND 'lease is **not** terminated' are 2 different outcomes

Neural Networks

- Currently only 1 estimator (multiple outputs not yet supported)
- Model trained on biased data involving **only** cases where tenant owed money
- Model is only used if SVM classifies the value as True
- Values: $[1, \infty]$

Machine Learning Limitations

- Predictions are only as good as the data it is trained on:
 - Dataset originates from Regie du Logement, appeals in higher courts not considered
 - Dataset only contains cases which involved a hearing
 - ~75% of dataset involves the landlord as plaintiff



Demo time!