ALGO MODEL COMMANDS for PRODIGY

~~prodigy terms.teach algo_terms en_core_web_lg --seeds "neural_network, decision_tree, random_forest" This didn't work well, gave random English words~~

prodigy sense2vec.teach algo_terms /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds "neural network"

prodigy terms.to-patterns algo_terms--label ALGO

{"label":"ALGO","pattern":[{"lower":"neural_network"}]}
{"label":"ALGO","pattern":[{"lower":"decision_tree"}]}
{"label":"ALGO","pattern":[{"lower":"random_forest"}]}
{"label":"ALGO","pattern":[{"lower":"neural_network|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"neural_networks|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"neural_net|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"genetic_algorithms|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"neural_nets|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"genetic_algorithm|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"evolutionary_algorithms|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"artificial_neural_networks|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"ANNs|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"deep_learning|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"backpropagation|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"natural_language_processing|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"Turing_machines|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"backprop|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"RNNs|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"reinforcement_learning|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"finite_state_machines|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"SVM|ORG"}]}
{"label":"ALGO","pattern":[{"lower":"turing_machine|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"NNs|NOUN"}]}

-COULD BE NEAREST NEIGHBOR OR NERUAL NETOWORK

{"label":"ALGO","pattern":[{"lower":"SVMs|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"gradient_descent|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"decision_trees|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"fuzzy_logic|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"FPGAs|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"logistic_regression|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"LSTM|ORG"}]}
{"label":"ALGO","pattern":[{"lower":"state_machines|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"RNN|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"MCMC|ORG"}]}

prodigy db-out algo_terms > /Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/algo_patterns.jsonl

Prodigy ner.manual algo_terms blank:en /Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/arxiv_train.jsonl --label ALGO

NOTES FOR ALGO LABELS:

Labeled	Did not Label	Rejected
Binary classification Nearest neighbor	NN (for nearest neighbor)	Machine learning Previous methods

Hidden Markov Models	Nearest neighbor (NN)	Additive model
Baldi_chauvin algorithm	classification	Efficient algorithm
Naïve Bayes models	KNN	Clustering algorithms
Mixture models	k-NN	Optimal algorithm
Logistic classification models	PCA	Existing methods
Logistic sequence prediction models	KNIFE	Our algorithm
Simulated annealing	PVM	Optimization technique
MCMC (exclude methiods)	LASSO	Resulting regression
Chinese Restaurant Process	Mutioutput kernel methods	Statistical model
Binary regression trees	Class probability estimation	Statistical methods
Random forests	Schoenberg transformations	Causal models
Support vector machines		The method/model
$ k$ nearest neighbors		
Classified constrained dimensionality reduction		
Latent dirichlet allocation		
Interaction component model		
K-nearest neighbors		
Manifold learning algorithm		
Manifold-based embedding algorithm		
k-nearest neighbors		
local linear embedding algorithm		
principle component analysis		
sparse linear model		
online boosting ~~algorithm~~		
Adaboost ~~algorithm~~		
Adaboost		
Online boosting		
Min-cut clustering		
Genetic algrithm		
Decision forest		
Autoencoder neural network		
Tree-based regressor		
Active set algorithm		
Boosting		
Tree ensembles		
Kernel ridge regression		
Kernel regression		
Latent variable models		
Network models		
Cellular automata model		
Fuzzy logic		
Binary classification		
k-means clustering algorithm		
k-means algorithm		
prototype vector machine		
kriging models		
conditional independence algorithms		
sparse PCA		
Prim’s algorithm		
Lloyd’s algorithm		
K- means		
Spectral clustering algrothims		

Dirichlet process mixtures of general linear models CART Bayesian trees Dirichlet process mixtures regression models Laplacian Support vector machines Expectation propagation Kernel based nearest neighbor approach Gradient descent Forward step-wise regression Slow feature analysis Expectationmaximization Latent variable model Bayesian canonical correlation Markov chain monte carlo Indian buffet process Kruskals algorithm Forests Gaussian graphical models Independent component analysis FastICA RobustICA Least-square support vector machine Ridge regression Bayesian network model Restricted boltzmn machines Deep belief networks Gaussian process latent variable model SVM Kernel induced random survival forests Random survival forests Bolzman machines Iterative detection-estimation Sparse linear regression Bagging Binary classifier Binary classification Single line search		

prodigy db-out algo_terms > /Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/algo_data_model.jsonl

python -m spacy pretrain /Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/arxiv_train.jsonl en_vectors_web_lg /Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/pretrain_algo_model/ --use-vectors

Prodigy ner.batch-train algo_terms en_vectors_web_lg --init-tok2vec /Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/pretrain_algo_model/model999.bin --output algo_terms_model --eval-split 0.2 --label ALGO

36.7%

Prodigy ner.make-gold algo_terms_correct ./tmp_model
/Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/arxiv_train.jsonl --label ALGO

Weird – this is the first model that didn't label anything!!!

36.7%

Prodigy ner.print-stream algo_terms
/Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/arxiv_train.jsonl --label ALGO

Prodigy ner.print-stream algo_terms_model
/Users/sashaqanderson/Dropbox/USGS/NER_Work/algo_model3/arxiv_train.jsonl --label ALGO