

MODEL 1: (GEO + ML Methods)

pip install sense2vec

```
prodigy sense2vec.teach terms_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"decision tree, decision trees, random forest, random forests, linear regression, regression, water balance, water
balance, aqueous model, subduction zone, species occurrence, transport model, sliding block, nonlinear
regression, chemical equilibrium, flow velocity, heat transport, linear sorption, Lagrangian transport, Hayes,
bayesian, differential equation"
```

```
prodigy sense2vec.teach terms_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"regression"
```

```
prodigy terms.to-patterns terms_method --label METHOD
```

```
prodigy db-out terms_method > /Users/sashaqanderson/Dropbox/USGS/NER_Work/method_patterns.jsonl
```

```
Prodigy ner.manual method_data blank:en /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl --
label METHOD
```

```
prodigy db-out method_data > /Users/sashaqanderson/Dropbox/USGS/method_data_model.jsonl
```

```
python -m spacy pretrain /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl en_vectors_web_lg
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/ --use-vectors
```

```
prodigy ner.batch-train method_data en_vectors_web_lg --init-tok2vec
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/model999.bin
```

```
Prodigy ner.batch-train method_data en_vectors_web_lg --init-tok2vec
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/model999.bin --output tmp_model --eval-
split 0.2 --label METHOD
```

40% Accuracy

```
Prodigy ner.make-gold method_data_correct ./tmp_model
/Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl --label METHOD
```

```
Prodigy ner.batch-train method_data_correct en_vectors_web_lg --init-tok2vec
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/model999.bin --output method_model --
eval-split 0.2 --label METHOD --n-iter 20
```

64% Accuracy

MODEL 2 (GEO + ML Methods):

pip install sense2vec

```
prodigy sense2vec.teach geo_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds "
heat_transfer"
```

prodigy sense2vec.teach geo_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds

"subduction_zone" good one!

prodigy sense2vec.teach geo_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds

"hydraulic " THOUGHT IT WAS CARS (BAD ONE)

prodigy sense2vec.teach geo_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds

"model" rejected most – just model words

prodigy sense2vec.teach geo_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds

"simulation" good one!

prodigy terms.to-patterns geo_method --label METHOD

```
{"label": "METHOD", "pattern": {"lower": "geologic|ADJ"}}
{"label": "METHOD", "pattern": {"lower": "plate_tectonics|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "weather_patterns|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "atmospheric_composition|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "tectonics|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "tectonic_movement|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "climate_system|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "Milankovitch_cycles|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "radioactive_decay|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "climate_model|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "climate_patterns|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "global_temperatures|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "greenhouse_effect|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "carbon_cycle|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "nuclear_decay|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "water_cycle|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "natural_language_processing|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "heat_transfer|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "thermal_conductivity|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "thermal_transfer|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "fluid_flow|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "heat_flow|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "mass_transfer|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "temperature_gradient|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "thermal_energy|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "evaporation|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "natural_convection|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "fluid_dynamics|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "energy_transfer|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "heat_capacity|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "boundary_layer|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "heat_energy|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "temperature_differential|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "thermal_mass|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "heat_dissipation|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "thermal_expansion|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "thermal_radiation|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "laminar_flow|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "latent_heat|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "fluid_mechanics|NOUN"}}
```

{"label": "METHOD", "pattern": {"lower": "critical_temperature|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "temperature_gradients|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "evaporation_rate|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "electrical_conductivity|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "atmospheric_pressure|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "energy_input|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "convection_currents|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "surface_tension|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "energy_loss|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "ionization|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "thermal_equilibrium|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "mass_flow_rate|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "flow_rate|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "pressure_gradient|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "exothermic_reaction|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "magnetic_field|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "differential_pressure|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "pressure_differentials|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "hydrostatic_pressure|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "subduction_zone|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "subduction_zones|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "subsidence|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "surface_water|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "subduction|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "water_table|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "aquifer|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "wave_action|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "sediments|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "ground_water|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "aquifers|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "liquefaction|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "water_tables|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "sedimentation|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "pollutants|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "water_flow|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "algal_blooms|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "waste_water|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "flooding|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "watersheds|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "water_pollution|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "geothermal_energy|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "wastewater|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "algae_blooms|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "soil_erosion|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "water_system|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "water_systems|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "hydraulic|ADJ"}}}

{"label": "METHOD", "pattern": {"lower": "hydraulic_system|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "model|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "simulation|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "computer_simulation|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "simulations|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "computer_simulations|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "neural_network|NOUN"}}}

{"label": "METHOD", "pattern": {"lower": "simulation_program|NOUN"}}}

```

{"label": "METHOD", "pattern": {"lower": "computer_model|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "neural_networks|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "genetic_algorithms|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "neural_net|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "neural_nets|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "evolutionary_algorithms|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "computational_models|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "genetic_algorithm|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "artificial_neural_networks|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "deep_learning|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "ANNs|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "computer_vision|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "Bayesian_inference|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "machine_vision|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "natural_language_processing|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "object_recognition|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "unsupervised_learning|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "lambda_calculus|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "NNs|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "reinforcement_learning|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "RNNs|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "signal_processing|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "SVMs|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "logistic_regression|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "LSTM|ORG"}}
{"label": "METHOD", "pattern": {"lower": "finite_element_analysis|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "FPGAs|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "supervised_learning|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "SVM|ORG"}}
{"label": "METHOD", "pattern": {"lower": "probability_theory|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "decision_trees|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "Bayesian|NORP"}}
{"label": "METHOD", "pattern": {"lower": "boolean_logic|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "graph_theory|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "Bayesian_methods|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "MCMC|ORG"}}
{"label": "METHOD", "pattern": {"lower": "dynamic_programming|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "Fourier_transforms|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "Fourier_analysis|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "linear_systems|NOUN"}}
{"label": "METHOD", "pattern": {"lower": "differential_equations|NOUN"}}

```

prodigy db-out geo_method > /Users/sashaqanderson/Dropbox/USGS/NER_Work/geomethod_patterns.jsonl

Prodigy ner.manual geomethod_data blank:en /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl -
-label METHOD

prodigy db-out geomethod_data > /Users/sashaqanderson/Dropbox/USGS/geomethod_data_model.jsonl

python -m spacy pretrain /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text_train40.jsonl
en_vectors_web_lg /Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model3/ **--use-vectors**

Prodigy ner.batch-train geomethod_data en_vectors_web_lg --init-tok2vec
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model3/model999.bin --output geo_model --eval-
split 0.2 --label METHOD

64% Accuracy

```
Prodigy ner.make-gold geomethod_data_correct ./geo_model
/Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text_train40.jsonl --label METHOD
```

```
Prodigy ner.batch-train geomethod_data_correct en_vectors_web_lg --init-tok2vec
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model3/model999.bin --output geo_model --eval-
split 0.2 --label METHOD --n-iter 20
```

68% Accuracy

MODEL 3 (GEO):

```
pip install sense2vec
```

```
prodigy sense2vec.teach geo3_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"water_flow"
```

```
prodigy sense2vec.teach geo3_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"heat_transfer"
```

```
prodigy sense2vec.teach geo3_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"subduction_zone"
```

```
prodigy sense2vec.teach geo3_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds "
mass_transfer"
```

```
prodigy sense2vec.teach geo3_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"soil_erosion"
```

```
prodigy sense2vec.teach geo3_method
/Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds "geothermal_energy" bad
one! Don't use again.
```

```
prodigy terms.to-patterns geo3_method --label METHOD
```

```
{"label": "METHOD", "pattern": [{"lower": "water_flow| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "reservoir| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "flow_rate| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "water_level| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "waterflow| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "water_table| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "water_movement| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "evaporation| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "drainage| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "reservoirs| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "surface_water| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "aquifer| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "sediment| NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "ground_water| NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_surface | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_levels | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "groundwater | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "aquifers | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ocean_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "seawater | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "waste_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sediments | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_column | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sea_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "salinity | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "fresh_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "pollutants | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "particulates | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "dissolved_oxygen | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_tables | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "salt_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "rain_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "evaporated_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "silt | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "contaminants | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_system | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "saltwater | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "compaction | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "rainfall | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "nutrient_solution | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "gas_exchange | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "particulate_matter | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "river_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "rainwater | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "surface_waters | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "wastewater | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "dissolved_gasses | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "underground_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_cycle | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "particulate | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "basins | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "oceans | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "snowmelt | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sea_floor | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "thermal_mass | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "source_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "temperature_gradient | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "basin | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "evaporation_rate | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "algae_blooms | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "phytoplankton | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "algal_blooms | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "freshwater | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "organic_matter | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ozone | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "waterways | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "lake_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "contaminates | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sedimentation | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sewage_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ocean_waters | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "seafloor | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "volcanic_ash | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ocean_floor | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_ways | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "eutrophication | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ocean_currents | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ocean_surface | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "saline_water | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "turbidity | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "greenhouse_gasses | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "topsoil | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "microorganisms | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "acidification | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "mass_transfer | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "heat_transfer | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "heat_exchange | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "fluid_flow | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "thermal_energy | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "Evaporation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "heat_equation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "Thermodynamics | ORG"}]}
{"label": "METHOD", "pattern": [{"lower": "finite_element_analysis | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "radioactive_decay | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "energy_transfer | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "gas_laws | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "heat_flow | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "thermal_radiation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "boundary_layer | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "ionization | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "atmospheric_pressure | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "temperature_gradients | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "laminar_flow | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "chemical_reactions | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "nuclear_decay | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "mass_flow_rate | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "magnetic_field | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "mass_flow | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "chemical_reaction | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "magnetic_fields | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "magnetic_flux | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "biological_systems | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "solar_radiation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "pressure_gradient | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "heat_radiation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "convection_currents | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "gravitational_fields | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "EM_radiation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "electromagnetic_radiation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "EM_waves | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "atmospheric_composition | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "upper_atmosphere | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "electromagnetic_waves | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "EM_fields | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "subduction_zone | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "Pacific_ocean | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "fault_lines | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "ice_sheets | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "fault_line | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "continental_shelf | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "coastal_areas | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "subduction_zones | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "jet_stream | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "subsidence | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "tectonic_plate | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "tectonic_plates | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "continental_crust | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "coastline | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "subduction | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "caldera | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "glaciers | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "coastlines | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "oceanic_crust | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "rising_seas | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "coastal_regions | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "gulf_stream | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "rising_sea_levels | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "volcanic_activity | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "subducted | VERB"}]}
{"label": "METHOD", "pattern": [{"lower": "wave_action | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "ice_shelf | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "rain_shadow | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "sea_level_rise | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "ocean | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "floodplain | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "coast_lines | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "volcanism | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "marshland | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "polar_regions | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "waterspouts | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "glacier | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "tectonic_activity | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "mountain_range | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "quakes | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "boreal_forests | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "sea_ice | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "wind_patterns | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "sea_level | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "sea_levels | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "volcanoes | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "mountain_ranges | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "shorelines | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "seabed | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "continental_plates | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "jet_streams | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "coast_line | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "plate_tectonics | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "tropics | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "polar_caps | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "deserts | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "lakes | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "temperate_zones | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "polar_ice_caps | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "lagoons | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "wetlands | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "weather_patterns | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "volcanic_eruptions | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "glaciation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "rivers | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "reefs | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "currents | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "dust_storms | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "lava_flow | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "polar_ice | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "earthquakes | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "grasslands | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "volcanic_eruption | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "volcanos | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "tectonic_movement | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "liquefaction | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "sea_life | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "hydrothermal_vents | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "coral_reefs | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "ice_caps | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "monsoons | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "tsunamis | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "vegetation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "precipitation | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "geysers | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "rain_forests | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "desertification | NOUN"}]}
{"label": "METHOD", "pattern": [{"lower": "climates | NOUN"}]}
```



```
{"label": "METHOD", "pattern": [{"lower": "rainforests | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sea-level | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ice_age | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ice_cap | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "rock_formations | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "snowpack | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ocean_levels | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "droughts | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "weather_systems | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "mountains | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "marshlands | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "hurricanes | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "seismic_activity | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "wind_currents | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "water_currents | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "marine_life | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ice_ages | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "rain_forest | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "icecaps | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "flash_floods | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "watersheds | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "dams | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "wildfires | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "marshes | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "icebergs | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "geothermal_heat | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "meteor_impacts | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "forests | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "eruptions | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sea_walls | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "air_currents | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "tropical_rainforest | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "rainforest | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "sea_bed | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "land_bridges | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "prairies | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ocean_acidification | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "geothermal_energy | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "soil_erosion | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "heat_transfer | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "thermal_transfer | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "mass_transfer | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "temperature_differential | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "surface_tension | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "pressure_differential | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "pressure_differentials | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "soil_erosion | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "deforestation | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "biodiversity | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "overfishing | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ecological_impact | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "erosion | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "pollution | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "overgrazing | VERB"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "reforestation | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "clear_cutting | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "acid_rain | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "carbon_sequestration | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "wildlife_habitat | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "fertilizers | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "crops | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "oil_spills | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "greenhouse_gases | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "ecosystems | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "agriculture | NOUN"}]}
```

```
{"label": "METHOD", "pattern": [{"lower": "green_house_gases | NOUN"}]}
```

```
{
  "label": "METHOD",
  "pattern": [{"lower": "habitats | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "food_supply | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "habitat | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "fossil_fuels | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "invasive_species | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "greenhouse_gas_emissions | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "air_pollution | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "CO2_emissions | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "fracking | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "native_species | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "fossil_fuel_emissions | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "carbon_emissions | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "biosphere | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "fossil_fuel | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "methane_emissions | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "green_house_gasses | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "overpopulation | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "greenhouse_emissions | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "carbon_emission | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "biomass | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "pollution | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "geothermal_energy | NOUN"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "geothermal | ADJ"}]
},
{
  "label": "METHOD",
  "pattern": [{"lower": "solar_energy | NOUN"}]
}
```

prodigy db-out geo3_method > /Users/sashaqanderson/Dropbox/USGS/NER_Work/geo3method_patterns.jsonl

Prodigy ner.manual geo3method_data blank:en

/Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text_train40.jsonl --label METHOD

prodigy db-out geo3method_data > /Users/sashaqanderson/Dropbox/USGS/geo3method_data_model.jsonl

python -m spacy pretrain /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text_train40.jsonl
en_vectors_web_lg /Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_geo3model/ --use-vectors

Prodigy ner.batch-train geo3method_data en_vectors_web_lg --init-tok2vec

/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_geo3model/model999.bin --output geo3_model --
eval-split 0.2 --label METHOD

53% Accuracy (oops – I didn't pretrain it first)

Prodigy ner.make-gold geo3method_data_correct ./geo3_model

/Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text_train40.jsonl --label GEO

Prodigy ner.batch-train geo3method_data_correct en_vectors_web_lg --init-tok2vec

/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_geo3model/model999.bin --output geo3_model --
eval-split 0.2 --label GEO --n-iter 20

69% (pretraining + make-gold)

Prodigy ner.make-gold geo3method_data_correct ./geo3_model

/Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl --label GEO

Ran into a problem with model 2 (NER labeled as METHOD and GEO)... start again.

Model 4

differential_equations
logistic_regression

pip install sense2vec

prodigy sense2vec.teach ml_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"differential equations, logistic_regression"
prodigy sense2vec.teach ml_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds
"simulation"

prodigy terms.to-patterns ml_method --label ML

```
{"label": "ML", "pattern": [{"lower": "differential_equations|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "logistic_regression|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "ODEs|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "differential_geometry|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "PDEs|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "linear_regression|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "topology|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "partial_differential_equations|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "Fourier_transforms|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "PDE|ORG"}]}
```

```
{"label": "ML", "pattern": [{"lower": "differential_equation|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "ANOVA|ORG"}]}
```

```
{"label": "ML", "pattern": [{"lower": "Fourier_analysis|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "linear_equations|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "Fourier_series|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "dynamic_programming|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "linear_transformations|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "ODE|ORG"}]}
```

```
{"label": "ML", "pattern": [{"lower": "genetic_algorithms|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "heat_equation|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "wave_equation|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "Laplace_transform|NOUN"}]}
```

```
{"label": "ML", "pattern": [{"lower": "Taylor_series|NOUN"}]}
```

```

{"label":"ML","pattern":[{"lower":"linear_functions|NOUN"}]}
{"label":"ML","pattern":[{"lower":"eigenfunctions|NOUN"}]}
{"label":"ML","pattern":[{"lower":"linear_approximation|NOUN"}]}
{"label":"ML","pattern":[{"lower":"simulation|NOUN"}]}
{"label":"ML","pattern":[{"lower":"neural_network|NOUN"}]}
{"label":"ML","pattern":[{"lower":"neural_networks|NOUN"}]}
{"label":"ML","pattern":[{"lower":"neural_net|NOUN"}]}
{"label":"ML","pattern":[{"lower":"neural_nets|NOUN"}]}
{"label":"ML","pattern":[{"lower":"genetic_algorithm|NOUN"}]}
{"label":"ML","pattern":[{"lower":"predictive_models|NOUN"}]}
{"label":"ML","pattern":[{"lower":"artificial_neural_networks|NOUN"}]}
{"label":"ML","pattern":[{"lower":"deep_learning|NOUN"}]}
{"label":"ML","pattern":[{"lower":"ANNs|NOUN"}]}
{"label":"ML","pattern":[{"lower":"computer_vision|NOUN"}]}
{"label":"ML","pattern":[{"lower":"machine_vision|NOUN"}]}
{"label":"ML","pattern":[{"lower":"natural_language_processing|NOUN"}]}
{"label":"ML","pattern":[{"lower":"statistical_models|NOUN"}]}
{"label":"ML","pattern":[{"lower":"reinforcement_learning|NOUN"}]}
{"label":"ML","pattern":[{"lower":"RNNs|NOUN"}]}
{"label":"ML","pattern":[{"lower":"NNs|NOUN"}]}
{"label":"ML","pattern":[{"lower":"SVMs|NOUN"}]}
{"label":"ML","pattern":[{"lower":"NLP|ORG"}]}
{"label":"ML","pattern":[{"lower":"SVM|ORG"}]}
{"label":"ML","pattern":[{"lower":"decision_trees|NOUN"}]}
{"label":"ML","pattern":[{"lower":"logistic_regression|NOUN"}]}
{"label":"ML","pattern":[{"lower":"Bayesian_inference|NOUN"}]}
{"label":"ML","pattern":[{"lower":"Bayesian_methods|NOUN"}]}

```

```
prodigy db-out ml_method > /Users/sashaqanderson/Dropbox/USGS/NER_Work/ml_patterns.jsonl
```

```
Prodigy ner.manual ml_data blank:en /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl --label ML
```

```
prodigy db-out ml_data > /Users/sashaqanderson/Dropbox/USGS/ml_data_model.jsonl
```

```
python -m spacy pretrain /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl en_vectors_web_lg  
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_mlmodel/ --use-vectors
```

```
Prodigy ner.batch-train ml_data en_vectors_web_lg --init-tok2vec  
/Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_mlmodel/model999.bin --output ml_model --eval-  
split 0.2 --label ML
```

0% Accuracy (Hmmm.. none in the actual training data?)

```
Prodigy ner.make-gold ml_data_correct ./ml_model  
/Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl --label ML
```

Model was so poor and with 0% accuracy, I decided to try another one instead of unclicking every wrong entity (just about every word/phrase)