



Disease Diagnosis by using NLP's Algorithms

- SDAIA BOOTCAMP T5
- PRESENTED BY:
MOHAMMED ALGHAMDI & HATIM ALSHEHRI

An orange brushstroke graphic that tapers from left to right, serving as a background for the title.

METHODOLOGY

- PROBLEM STATEMENT
- TOOLS
- DATA STORY
- DATA PRE-PROCESSING
- TOPIC MODELING
- SUPERVISED LEARNING
- DEMO
- FUTURE WORK



Problem Statement

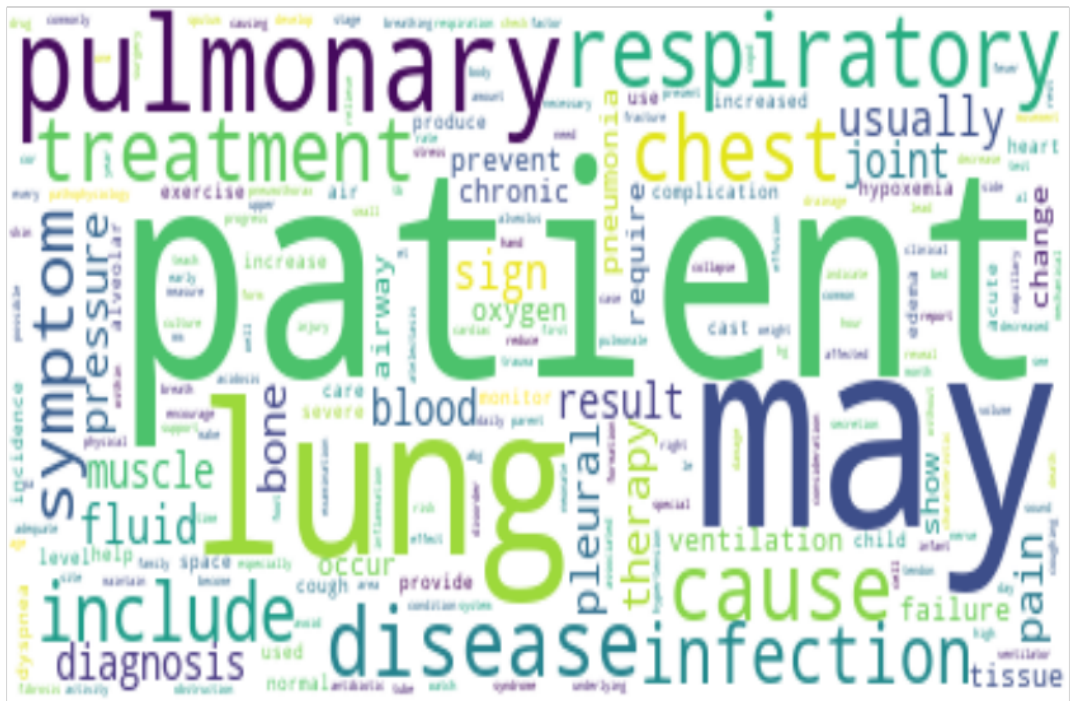
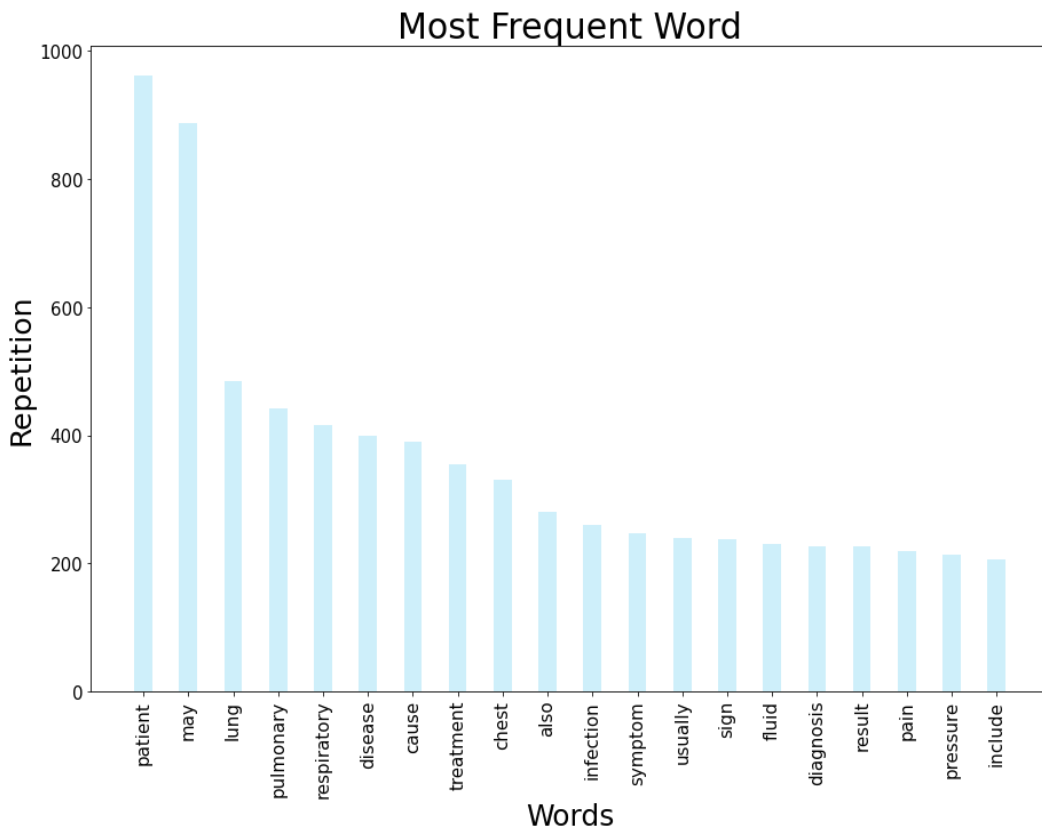
- The medical field has become more sophisticated and advanced than it was before, but early diagnosis of diseases helps to reduce the consequences of complications, that's thing inspired us and we recognized some people haven't the ability to know what type of diseases they have, especially at an early stage of diseases.



TOOLS

- Numpy
- Pandas
- sklearn
- PyPDF2
- gensim
- nltk
- matplotlib
- SpaCy
- Flask

DATA ST



DATA PRE-PROCESSING

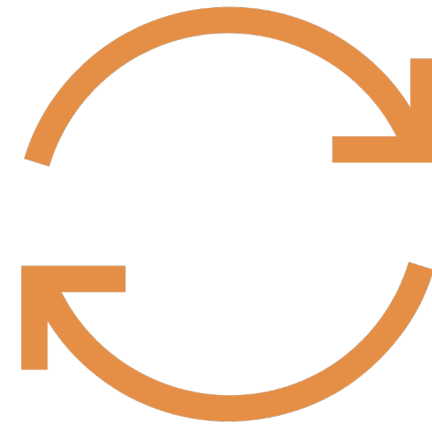
- Remove any digits and expression
- Remove entities with SpaCy
- Remove stop words in English & topic domain
- Lemmatization
- Vectorization

English Stop-words: 179

POS(Adjective): 1452

Topic Domain: 216

Total: 1847



Sample of Removail Words

Top 10 common words in Buseinsses Domain:

patient: 961

May: 887

lung: 484

Pulmonary: 442

respiratory: 417

Disease: 400

Cause: 390

Treatment: 355

Chest: 330

Also: 280

- 'patient', 'serosanguineous', 'last', 'bullous', 'gonadal', 'alkaline', 'uncomplicated', 'comprehensive', 'neurovascular', 'tear', 'neoplastic', 'hypnotic', 'markedly', 'bedcover', 'unavailable', 'natriuretic', 'ligamentous', 'permissible', 'stable', 'midaxillary', 'minor', 'difficult', 'cervical', 'variable', 'honeycombing', 'subcutaneous', 'extruded', 'inhaled', 'enteropathy', 'direct', 'bedridden', 'bronchopleural', 'brittle', 'fetal', 'gastric', 'edematous', 'pelvis', 'myoneural', 'mattress', 'milky', 'heberden', 'lobectomy', 'bright', 'unaffected', 'reflex', 'articular', 'nonproductive', 'sign', 'uricosuric', 'pollutant', 'outdoor', 'anesthesiologist', 'adverse', 'anorexia', 'definite', 'supersaturated', 'past', 'persistent', 'noxious', 'crackle', 'femur', 'antispasmodic', 'airborne', 'bacterial', 'annual', 'indistinct', 'mean', 'asterixis', 'fresh', 'croupy', 'obstructive', 'free', 'mosaic', 'malleable', 'deep', 'removable', 'noticeable', 'lymphoid', 'arteriosclerotic', 'closed', 'petal', 'previous', 'corrective', 'gray', 'malignant', 'uncommon', 'sternal', 'flat', 'firstborn', 'structural', 'palsy', 'interior', 'gelatinous', 'idiosyncrasy', 'neural', 'basilar', 'homelessness', 'bottom', 'confirm', 'static', 'endocarditis', 'applicable', 'intensive', 'alert', 'spinal', 'meniscal', 'special', 'tophus', 'copious', 'stressful', 'padded', 'clear', 'wrong', 'connective', 'cachectic', 'cold', 'faisy', 'aged', 'membranous', 'poisonous', 'frontal', 'lumbar', 'mediastinum', 'slow', 'standard', 'joint', 'human', 'nasal', 'atypical', 'fifth', 'better', 'weekly', 'third', 'dextrose', 'skillful', 'central', 'extracellular', 'thorough', 'record', 'general', 'subside', 'conventional', 'symptom', 'neuromuscular', 'tenacious', 'paradoxical', 'eosinophil', 'consequent', 'fluorescent', 'mucous', 'whole', 'prevalent', 'improve', 'peroneal', 'nitrous', 'hemolytic', 'hypoxemic', 'radial', 'lymphadenopathy', 'unaware', 'referral', 'continuous', 'respiratory', 'neonate', 'linear', 'slight', 'longer', 'residual', 'electrical', 'certain', 'amytrophic', 'orthopaedic', 'inverted', 'instrumental', 'cellular', 'outer', 'overall', 'mandatory', 'cytotoxic', 'ill', 'reserved', 'cooperative', 'hypoalbuminemia', 'intractable', 'shunting', 'elderly', 'nutritional', 'overmedicate', 'nonsurgical', 'tender', 'ventricular', 'maximal', 'peptic', 'transudative', 'hot', 'subperiosteal', 'buttock', 'neurogenic', 'parentæinfant', 'cerebrospinal', 'immunosuppressive', 'saccular', 'mature', 'incomplete', 'neurologic', 'lobular', 'strongest', 'apneic', 'washed', 'subtrochanteric', 'inner', 'substernal', 'microbiological', 'adjacent', 'hyperresonant', 'inotropic', 'front', 'neuroarthropathy', 'radiographic', 'easier', 'bizarre', 'capillary', 'initial', 'posterior', 'mental', 'carbonic', 'underdeveloped', 'fewer', 'aerobic', 'alternate', 'main', 'absent', 'heavy', 'limp', 'yellow', 'medial', 'uninterrupted', 'cranial', 'patellar', 'familial', 'steroid', 'kind', 'upper', 'ricketsial', 'pickwickian', 'overwhelmed', 'recessive', 'narrow', 'destructive', 'positive', 'impulse', 'higher', 'asynchrony', 'basic', 'primary', 'lymphoproliferative', 'loud', 'spastic', 'fluid', 'vertebral', 'epiglottal', 'ancillary', 'embolic', 'compensatory', 'refractory', 'german', 'anticoagulant', 'uncontrolled', 'prescribed', 'osmotic', 'condensate', 'particular', 'light', 'painless', 'gold', 'capacitynormal', 'external', 'sure', 'depressionhead', 'probable', 'close', 'hollow', 'atherosclerotic', 'differential', 'exaggerated', 'continual', 'irritant', 'run', 'detailed', 'immediate', 'cyanotic', 'facial', 'ongoing', 'similar', 'immature', 'live', 'seventh', 'undetermined', 'subtle', 'unexplained', 'relative', 'autoimmune', 'atrophic', 'steady', 'turkish', 'multiplied', 'mixed', 'mucopurulent', 'paralyzed', 'underway', 'urinary', 'spasmodic', 'anginal', 'entrapment', 'proliferative', 'shorter', 'partial', 'nontraumatic', 'careful', 'ineffective', 'femoral', 'uptake', 'exact', 'classic', 'prolonged', 'foreign', 'histamine', 'hypertensive', 'soft', 'mycoplasmal', 'characteristic', 'precordial', 'multifactorial', 'squamous', 'reticular', 'pneumomediastinum', 'neoplasm', 'consistent', 'adequate', 'excessive', 'middle', 'mobile', 'retrospective', 'outgrow', 'systemic', 'second', 'turbulent', 'reduced', 'prospective', 'complicated', 'intermediate', 'terminate', 'identifiable', 'patient', 'may', 'disease', 'cause', 'treatment', 'also', 'symptom', 'usually', 'sign', 'diagnosis', 'result', 'pain', 'include', 'pressure', 'lung', 'pulmonary', 'respiratory', 'chest', 'fluid', 'complication', 'change', 'blood', 'infection', 'therapy', 'prevent', 'acute', 'care', 'child', 'level', 'air', 'use', 'severe', 'help', 'used', 'exercise', 'normal', 'incidence', 'pneumonia', 'tissue', 'show', 'chronic', 'failure', 'cast', 'increased', 'monitor', 'hypoxemia', 'produce', 'edema', 'increase', 'space', 'occurs', 'cough', 'alveolar', 'heart', 'pathophysiology', 'sputum', 'provide', 'decreased', 'pneumothorax', 'test', 'special', 'tube', 'condition', 'common', 'surgery', 'secretion', 'fibrosis', 'disorder', 'pa', 'area', 'form', 'cell', 'skin', 'drainage', 'tb', 'year', 'commonly', 'check', 'teach', 'rest', 'watch', 'encourage', 'underlying', 'consideration', 'et', 'early', 'hour', 'family', 'need', 'effusion', 'body', 'drug', 'support', 'rate', 'syndrome', 'requires', 'inflammation', 'abg', 'side', 'infant', 'however', 'upper', 'cor', 'pulmonale', 'ventilator', 'mechanical', 'breath', 'maintain', 'foot', 'day', 'bed', 'parent', 'especially', 'fever', 'culture', 'system', 'within', 'factor', 'amount', 'death', 'movement', 'progress', 'volume', 'one', 'stage', 'report', 'avoid', 'respiration', 'trauma', 'occur', 'atelectasis', 'hand', 'includes', 'weight', 'tendon', 'hypertension', 'le', 'time', 'lead', 'damage', 'causing', 'require', 'activity', 'injury', 'risk', 'mm', 'measure', 'examination', 'nerve', 'stress', 'make', 'al', 'see', 'decrease', 'age', 'hgcage', 'month', 'coughing', 'develops', 'formation', 'without', 'site', 'every', 'reduce', 'relieve', 'effect', 'percussion', 'ordered', 'develop', 'affect', 'loss', 'flow', 'lesion', 'technique', 'exposure', 'gas', 'finding', 'procedure', 'begin', 'wall', 'immediately', 'type', 'response', 'position', 'needed', 'administer', 'control', 'ass', 'increasing', 'although', 'tell', 'output', 'give', 'analysis', 'history', 'often', 'week', 'home', 'perform', 'function', 'typically', 'frequently', 'adult', 'indicate', 'administration', 'explain', 'using', 'suggest', 'called', 'center', 'head', 'people', 'resulting', 'including', 'period', 'feature'

First Step : determine body systems using TOPIC MODELING

CorEx

LSA

LDA





CorEx

- **Topic 0:**
staphylococci, fluorescent, metatarsal, accompanied, permitted, fraction, bloodstream, digoxin, rsv, elderly
- **Topic 1:**
acetaminophen, vitamin, generally, methacholine, local, advise, alive, trimethoprim, booming, deliver
- **Topic 2:**
gold, vendor, aminotransferase, indication, ptt, referred, cease, globally, intravascular, meninges

LDA

- **Topic 0:**

al, ventilation, journal, retrieved, http, medicine, hypertension, injury, antibiotic, bronchiectasis

- **Topic 1:**

pleural, oxygen, copd, american, report, case, clinical, death ,pediatric, tuberculosis

- **Topic 2:**

airway, dyspnea, silicosis, technique, cardiac, muscle, atelectasis, abscess, massive



LSA

- **Topic 0:**

bone, muscle, ear, otitis, hearing, membrane, bleeding, airway, deformity, obstruction

- **Topic 1:**

ear, otitis, hearing, throat, sinusitis, bleeding, nose, membrane, externa, obstruction

- **Topic 2:**

ventilation, oxygen, airway, copd, breathing, acidosis, hg, bronchiectasis, artery, collapse,

The background features a large, abstract orange brushstroke that starts from the top left and sweeps across the upper right portion of the slide. A smaller, similar brushstroke is located in the bottom left corner.

Best Topic Model: LSA

BASED ON THE PREVIOUS
RESULT WE FOUND THE
BEST TOPIC MODEL IS LSA

A large, irregular orange brushstroke shape serves as a background for the text, centered on a white page. The brushstroke has a textured, hand-painted appearance with varying shades of orange and some darker edges.

Cosine Similarity

WE USED COSINE SIMILARITY TO FIND THE
USER INPUT SIMILARITY WITH THE MOST
RELATIVE BODY ORGAN SYSTEM BASED
ON TOPIC MODELING.

Second Step : Get symptoms of a disease

An orange brushstroke graphic with a rough, hand-painted edge, containing the text "SUPERVISED LEARNING" in white capital letters.

SUPERVISED
LEARNING

- LogisticRegression
- accuracy score: 76



Demo



HERE IS A WORK
DEMO EXECUTED
LOCALLY WITH FLASK

FUTURE WORK

- Improving model performance
- Increase the data source
- Implementing language correction
- leveraging the conversation between the doctor and patient, in order to build an integral NLP model



Data sources

- Professional Guide to Diseases 11th edition
- www.cdc.gov
- www.webmd.com

Thank you

