



Identification of Intra-Domain Ambiguity using Transformer-based Machine Learning

Ambarish Moharil

Dept. of Instrumentation & Control Engineering

Vishwakarma Institute of Technology

Pune, India

Email: ambarish.m02@gmail.com

Arpit Sharma

Dept. of Electrical Engineering & Computer Science

Indian Institute of Science Education and Research

Bhopal, India

Email: arpit@iiserb.ac.in

Experimentation Journal for the C'_{MD} Corpus (Mixed Corpus)



Table of Contents

Content	Page Number
Section 1 :- Experiments from the paper	Pg 3-37
State	Pg 3-7
Bug	Pg 8-11
Concrete	Pg 12-15
Mobile	Pg 16-19
Plant	Pg 19-23
Drive	Pg 24-30
Bank	Pg 30-37
Section 2 :- Some more words	Pg 37-59
Application	Pg 37-42
Function	Pg 42-46
Language	Pg 46-50
Float	Pg 50-55
Stack	Pg 56-59
<u>Section 3:- Overlapping Clusters</u>	Pg 60-72
Opaque	Pg 60-63
Backup	Pg 64-68
Fork	Pg 68-72



Multi Domain Analysis

This report intends to present the results of our experimentations on the C'_{MD} corpus. It is to be noted that after analyzing the cosine similarity plots (threshold plots) for all the target words selected for our experimentation, we decided to keep the threshold for aggregating relevant context words for each respective cluster as 0.45. The value 0.45 was calculated by taking the average of all the threshold values computed for the C'_{CS} corpus. As the threshold plots of both the clusters show a significant amount of similarity, we lock the threshold for collecting the relevant context words at 0.45 in this series of experiments on the C'_{MD} corpus.

Section 1:- Experiments From The Paper

Target word: - State

No of clusters: - 2

Deterministic Method :- Elbow Method

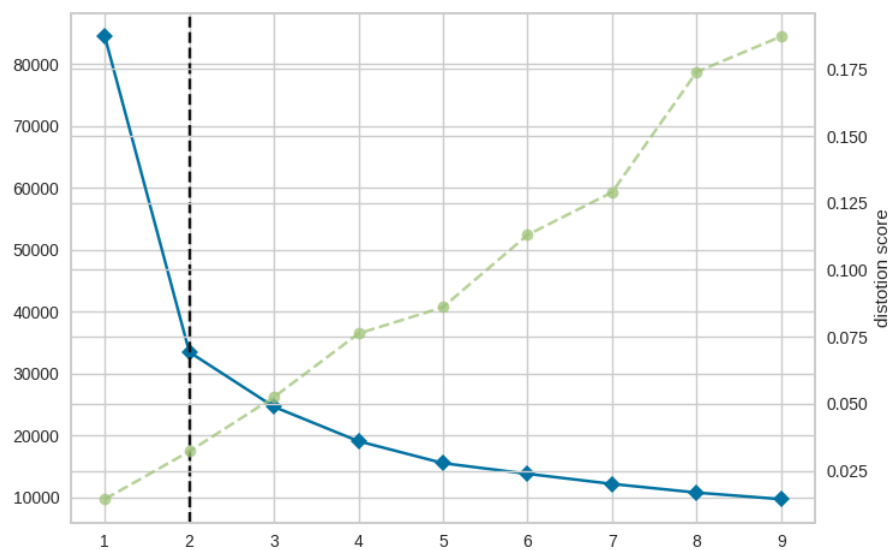
Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

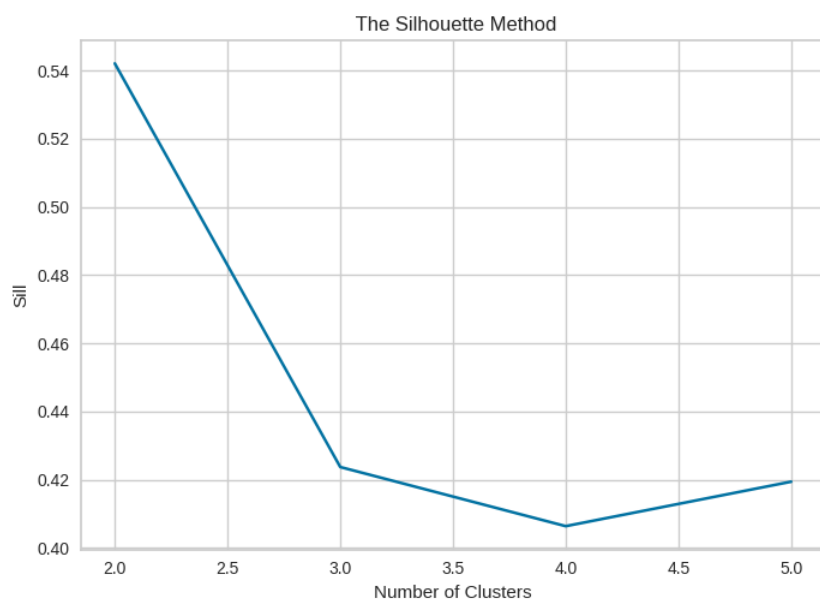
YES

Elbow = Silhouette = K = 2

K_2 Silhouette score = 0.54



Elbow Plot for the word State



Silhouette Plot for the word State

Cluster 0: -

Implied Meaning: - “Condition”, “Situation” (This cluster implies the same context as seen in the Cluster 0 of CS Domain.

Example Sentences: -

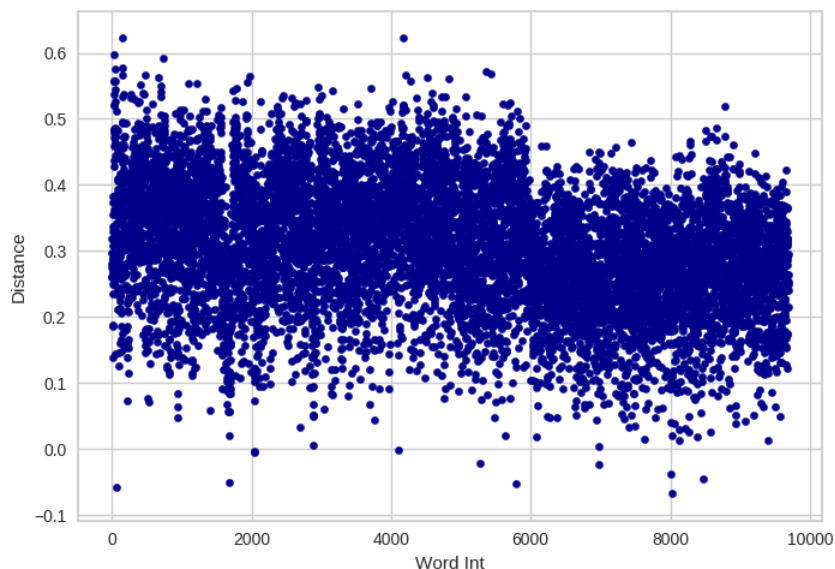
1) Instead of create (), a functional-style definition of an abstract stack may assume the existence of a special stack state, the empty stack, designated by a special symbol like Λ or "(); or define a bottom () operation that takes no arguments and returns this special **stack state**.

2) Starting from an **initial state** and initial input (perhaps empty),[6] the instructions describe a computation that, when executed, proceeds through a finite[7] number of well-defined successive states, eventually producing "output"[8] and terminating at a final ending state.

3) The **simplest state** due Post Turing says essentially effective method solving certain sets problems exists one build machine solve problem set human intervention beyond inserting question later reading answer. All three definitions are equivalent, doesn't matter one used.

4) The simplest of **these to state** (due to Post and Turing) says essentially that an effective method of solving certain sets of problems exists if one can build a machine which will then solve any problem of the set with no human intervention beyond inserting the question and (later) reading the answer.

Context Words: - Behaviour, trajectory, content, configuration, conditions, shape, behave, current, trace, pattern, representations, uncertainty, arrangements etc.



Threshold Scatter plot for a label in cluster 0

Cluster 1 : -

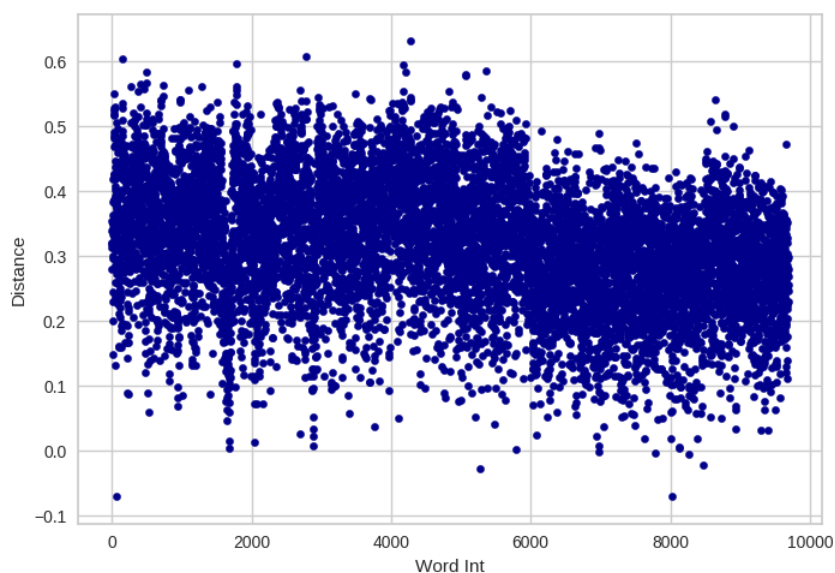
Implied Meaning: - “A geographical State” (A very peculiar context has been detected by the model which wasn’t present in the CS domain) The cluster 1 of CS domain becomes redundant in this case. Either it is merged with the first cluster or is not discerned properly by the algorithm. Not bad in either of the case.

Example Sentences: -

- 1) The tanks were assigned to purpose-created E Squadron of the Rhodesian Armoured Corps. It is unknown whether these tanks were passed on to the successor **state** or destroyed during the Rhodesian Bush War. Russia – At least 3,000 inherited from the Soviet Union.
- 2) 600 T-55s were ordered in 1973 from the Soviet Union and delivered between 1973 and 1975 aid. Passed on to the **successor state**. Poland – 3,000 T-54, T-54A, T-54AD and T-54AM produced between 1956 and 1964.
- 3) The College was chartered on January 1, 2006 under a reorganization. The College includes two schools: the Mewbourne School of Petroleum and Geological Engineering and the ConocoPhillips School of Geology and Geophysics, and the Oklahoma Geological **Survey a state** agency mandated by the Oklahoma Constitution. The main offices for the College are located in Sarkeys Energy Center on the northeast corner of campus.

4) It is 8 km from Gujarat **state capital** Gandhinagar, and 23 km from Gujarat's largest city Ahmedabad. PDPU has been awarded Scientific & Industrial Research Organization SIRO recognition by Department of Scientific and Industrial Research, Ministry of Science & Technology, Government of India. The university has four schools, located on the same campus.

Context Words: - municipal, statewide, jurisdiction, provincial, legislative, province, county, domestic, nation, interstate, local, tribal, county, commonwealth, societal, regional etc.



Threshold plot for a label in cluster 1

Target word: - Bug

No of clusters: - 2

Deterministic Method :- C-Elbow Method

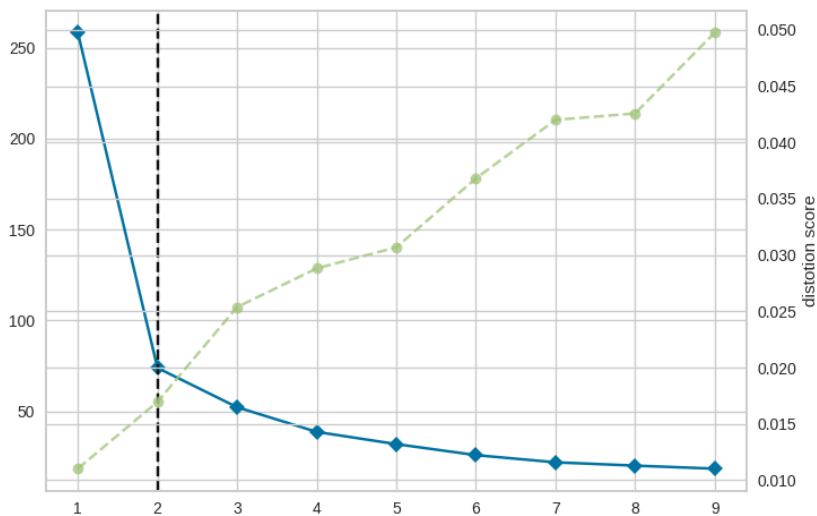
Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

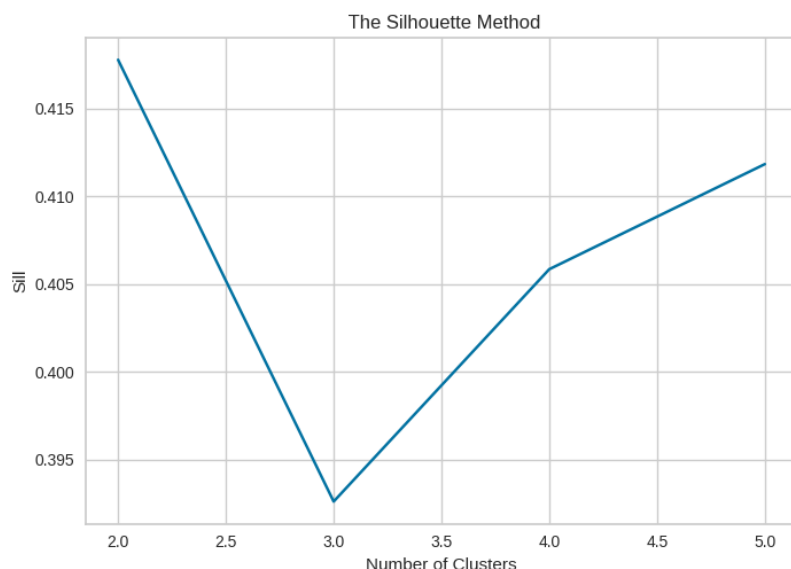
YES

No of clusters

$K = 2$



Elbow Graph for Bug



Silhouette Method Graph for Bug

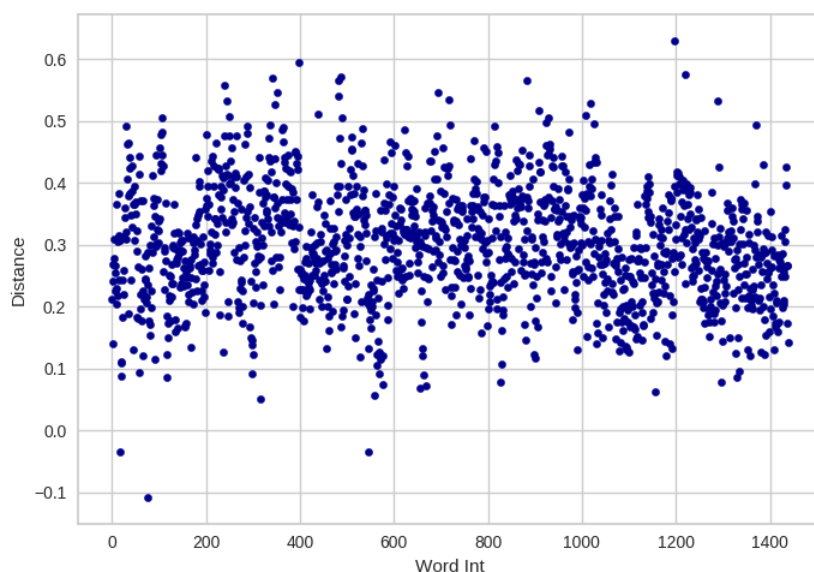
Cluster 0

Implied Meaning: - “Error” (This was the only context that was extracted in the CS Domain)

Example Sentences: -

- 1) Most **bugs** arise from mistakes and errors made in either a program's design or its source code, or in components and operating systems used by such programs.
- 2) Since the early minicomputers used magnetic core memory, which did not lose its information when power was off, these bootstrap loaders would remain in place unless they were erased. Erasure sometimes happened accidentally when a **program bug** caused a loop that overwrote all of memory.
- 3) **Software bug**. The actual first computer bug, a moth found trapped on a relay of the Harvard Mark II computer. Errors in computer programs are called "bugs".
- 4) But in some cases, they may cause the program or the entire system to “hang,” becoming unresponsive to input such as mouse clicks or keystrokes, to completely fail, or to crash. Otherwise benign bugs may sometimes be harnessed for malicious intent by an unscrupulous user writing an exploit, code designed to take advantage of a **bug** and disrupt a computer's proper execution.

Context Words: - *error, problems, failure, mistakes, flaw, flaws, stuck, invalid, anomaly, issue, errors, loop, feature, defect, disrupt etc.*



Threshold Scatter plot for a label in cluster 0

Cluster 1

Implied Meaning: - “An insect” (This a new context of the word that has been captured in the mixed corpus)

Example :-

1) Application methods for household insecticides

Pest management in the home begins with restricting the availability to insects of three vital commodities: shelter, water and food. If insects become a problem despite such measures, it may become necessary to control them using chemical

methods, targeting the active ingredient to the particular pest. Insect repellent, referred to as "bug spray", comes in a plastic bottle or aerosol can. Applied to clothing, arms, legs, and other extremities, the use of these products will tend to ward off nearby insects.

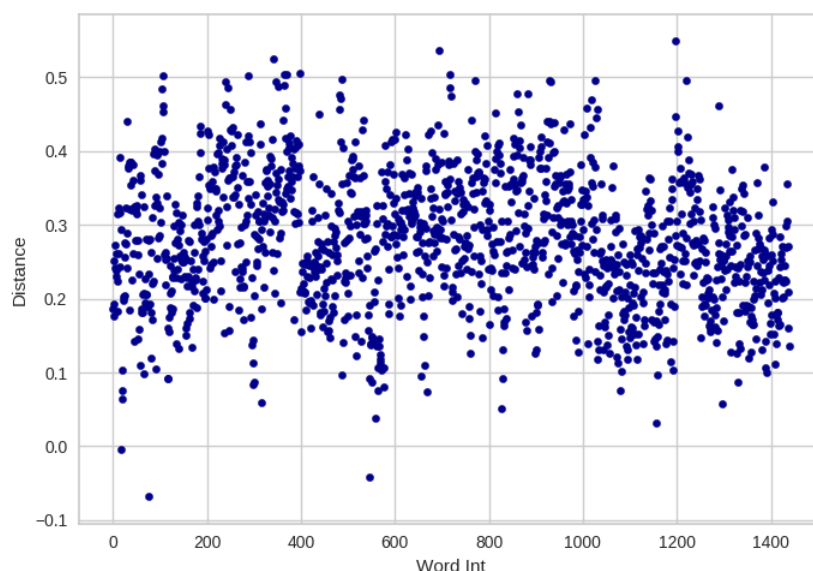
2) An insect hotel, also known as a bug hotel or insect house, is a manmade structure created to provide shelter for insects. They can come in a variety of shapes and sizes depending on the specific purpose or specific insect it is catered to.

3) Medical-grade diatomite has been studied for its efficacy as a deworming agent in cattle; in both studies cited the groups being treated with diatomaceous earth did not fare any better than

control groups. It is commonly used in lieu of boric acid, and can be used to help control and possibly eliminate bed bug, house dust mite, cockroach, ant and flea infestations

4) DEET a bug repellent, and caffeine were found in all water types and virtually in all samples. Triclosan in anti-bacterial soap & toothpaste was found in all water types, but detected in higher levels parts per trillion in reclaimed water than in surface or groundwater.

Context Word:- worm, flea, insect, cockroach, rats, pigeons, bacterial, dee, worm, pit, ant etc.



Threshold Scatter Plot for a label in cluster 1

Target word: - Concrete

No of clusters: - 2

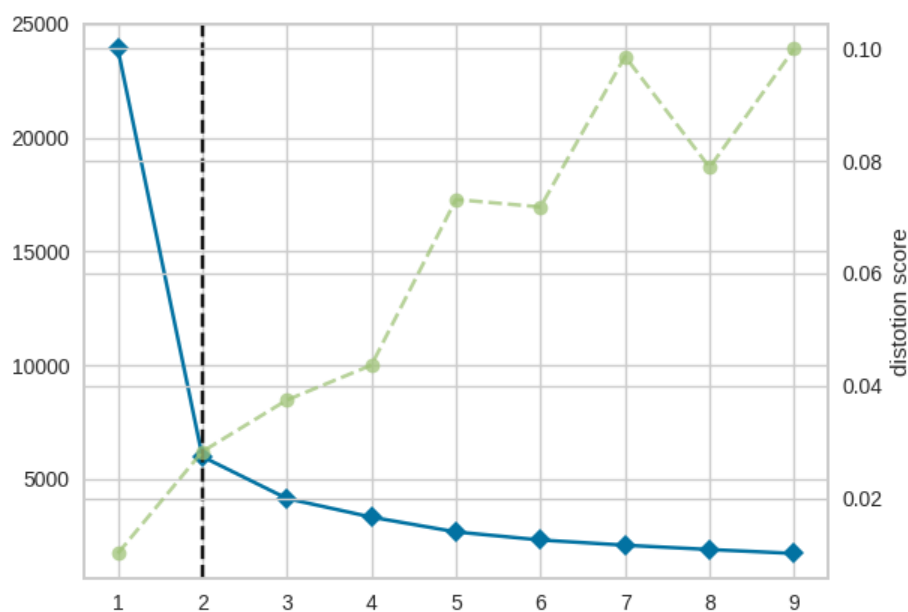
Deterministic Method :- Elbow Method

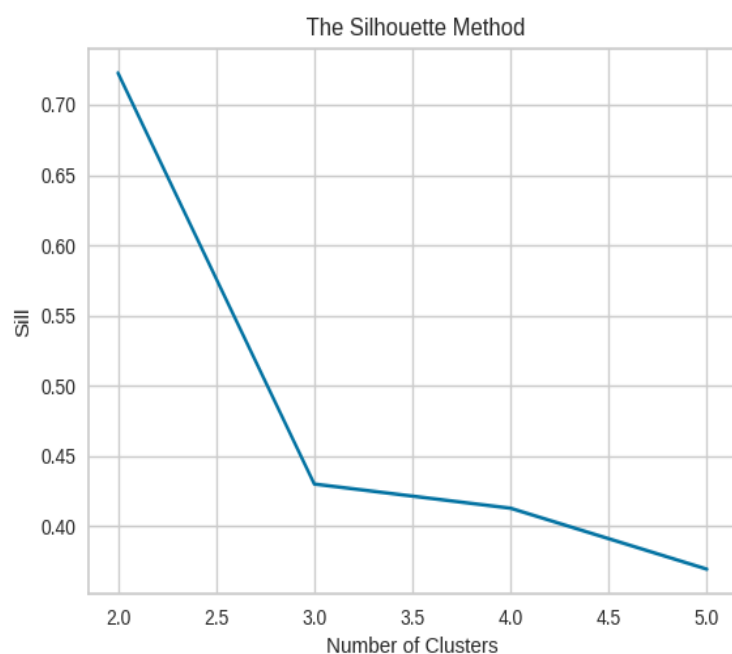
Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

YES

$K = 2$





Cluster 0

Implied Meaning : - “not abstract or imaginary, well established.”

Example Sentences : -

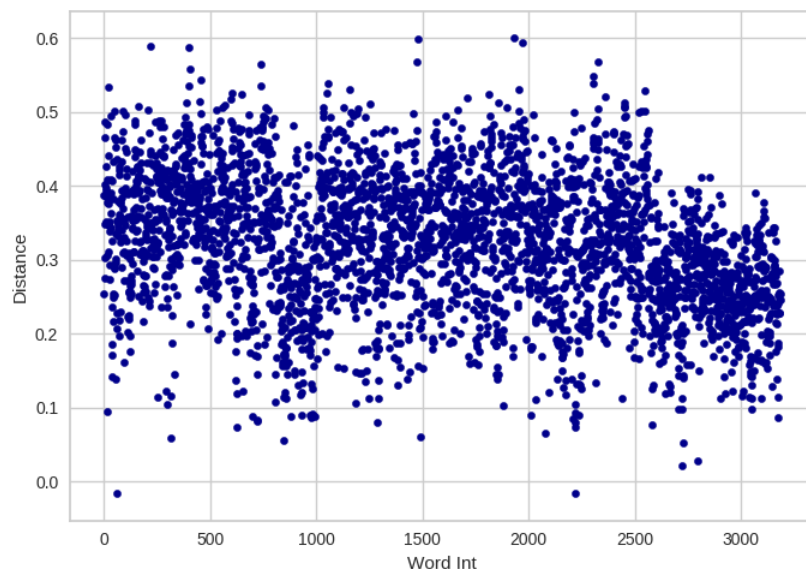
1) In a functional-style definition there is no need for a create operation. Indeed, there is no notion of "stack instance". The stack states can be thought of as being potential states of a singlestack structure, and two-stack states that contain the same values in the same order are considered to be identical states. This view actually mirrors the behavior of some **concrete** implementations, such as linked lists with hash cons.

2) The implementation of the module—namely, the bodies of the procedures and the **concrete** data structure used—can then be hidden from most clients of the module. This makes it possible to change the implementation without affecting the clients.

3) Many of the above **concrete notions** can be reinterpreted in this light, for example, the transpose matrix AT describes the transpose of the linear map given by A , with respect to the dual bases.

4) If certain algebraic objects (say, rational or integer solutions to certain equations) can be shown to be in the tail of certain sensibly defined distributions, it follows that there must be few of them; this is a very **concrete non-probabilistic** statement following from a probabilistic one.

Context Words : - basic, broad, complex, reference, understood, tangible, practical, valuable, feasible, valuable, suitable, reinforcement, relevant, precise, logical, rigid, detailed, substantiate, explain, defining, irrelevant, exhibit, composite, established etc.



Threshold Plot for cluster 0

Cluster 1

Implied meaning: - “a strong hard building material composed of sand and gravel and cement and water”

Example Sentences: -

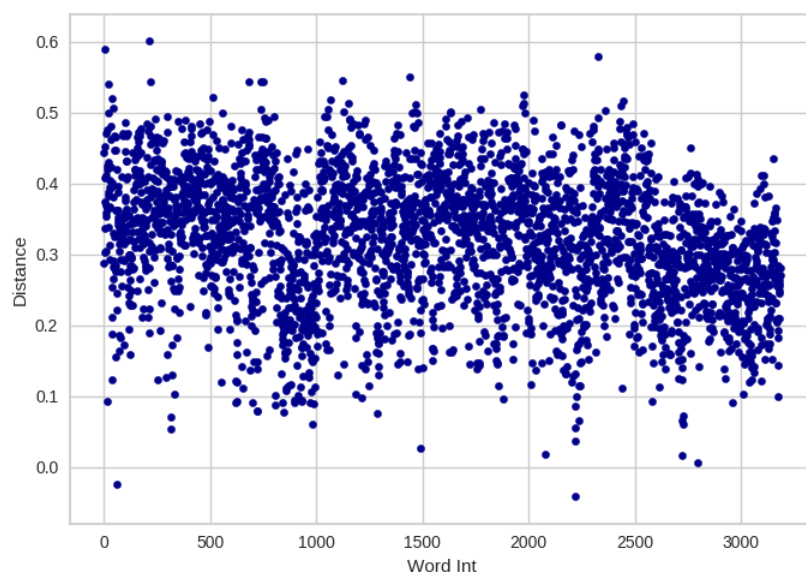
1) The three surviving concrete shafts are thought to be the first pre-cast reinforced concrete structures in Queensland, and examples of the earliest application of the Monier system of reinforced concrete construction in Queensland. The place demonstrates rare, uncommon or endangered aspects of Queensland's cultural heritage. They provide rare surviving evidence of this early use of true reinforced concrete, and are significant as indicators of the technically advanced state of municipal engineering and construction in Brisbane at the turn of the 20th century.

2) These structures weakened over their lifespan leading to increased risk of collapse. Newer cesspools are cast concrete which dramatically decrease any risk of collapse. All new construction in areas without waste sewer systems use these new precast cesspools.

3) Condrain conducts business mainly within the Greater Toronto Area, the Golden Horseshoe, Western Ontario, the Ottawa Region as well as in other provinces and the United States. Companies in the Condrain Group construct roads; supply recycled and virgin aggregate materials; produce **concrete pipe**, precast units and lighting poles; manufacture valves and fittings; install gas lines, power lines, fibre optic cables, and street lighting; build houses and commercial and industrial buildings; and develop large tracts of land for community living.

4) Ontario **Concrete** Products OCP began operations 2002 as a **precast concrete pipe** manufacturing company using robotic technology. == References == The Crossness Sewage Treatment Works at Crossness TQ487805 in the London Borough of Bexley is located at the eastern end of the Southern Outfall Sewer.

Context Words:- sandstone, steel, stucco, pillars, substance, waters, graphite, glass, arches, tiles, cement, boulders, rock, slate, basalt, masonry, foundations, coping, brick, slab, strengthen, walls, rocks, asphalt, quarry, welding, beams, embankment, bridges etc.



Threshold plot for cluster 1

Target word: - Mobile

No of clusters: - 2

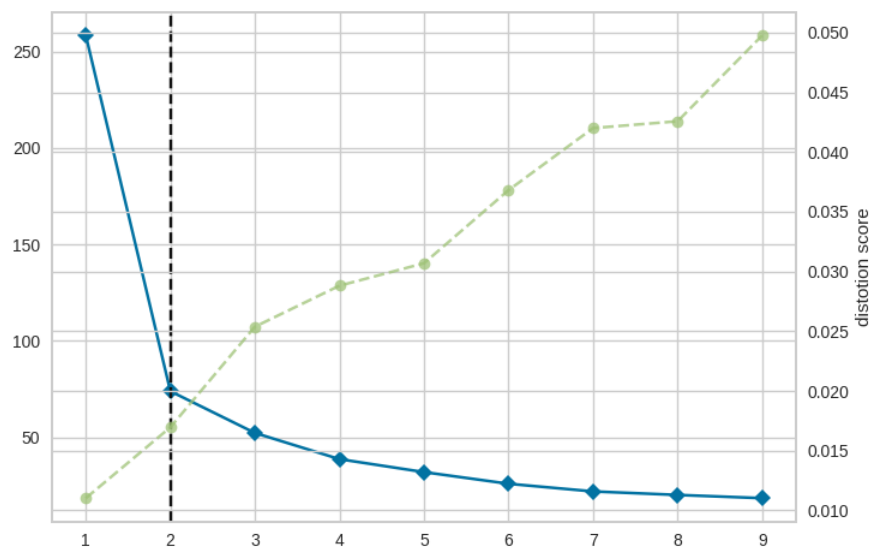
Deterministic Method :- Elbow Method

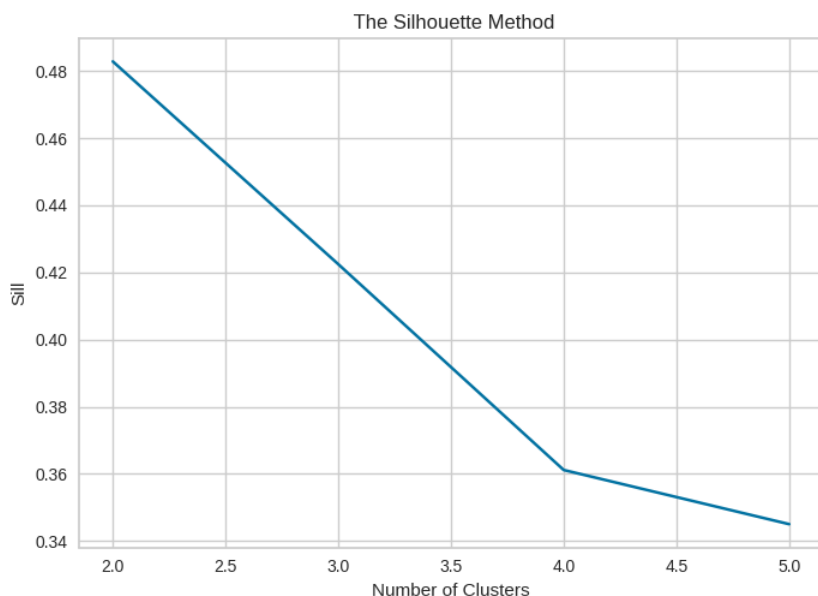
Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

YES

$K=2$





Cluster 0

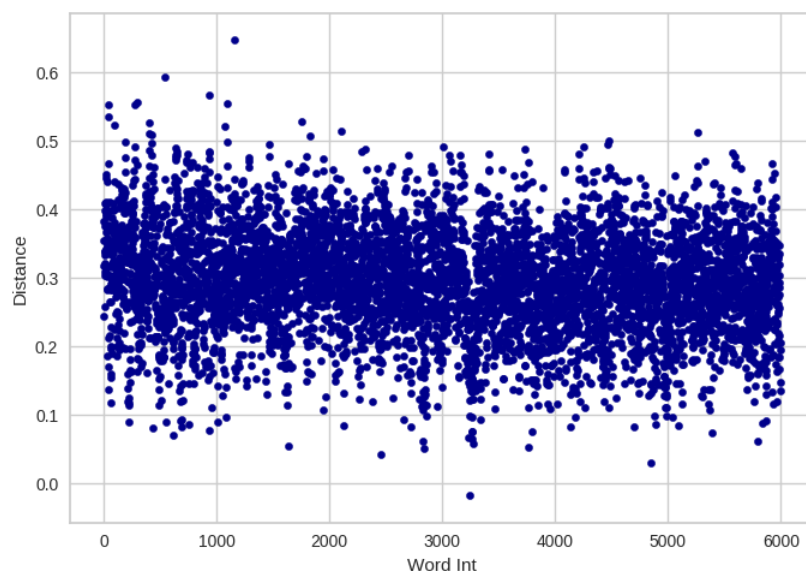
Implied Meaning: - “A cell phone”

Example Sentences: -

- 1) The development of the tablet computer was enabled by several key technological advances. The rapid scaling and miniaturization of MOSFET transistor technology (Moore's law), the basic building block of **mobile devices** and computing devices
- 2) Specific examples include speech coding and transmission in digital **mobile phones**, room correction of sound in hi-fi and sound reinforcement applications, analysis and control of industrial processes, medical imaging such as CAT scans and MRI, audio crossovers and equalization, digital synthesizers, and audio effects units.[18]
- 3) This process is, however, usually quite inexpensive, and commonly used. "Wizard of Oz": This is a comparatively uncommon technique but has seen some use in **mobile devices**.
- 4) Later studies have used chestnuts, oak acorns, and Moringa oleifera drumstick tree for the same purpose. Other research has examined the use of doped semiconductors to increase the production of oxygen radicals under solar UV-A.

Recently, researchers at the National Centre for Sensor Research and the Biomedical Diagnostics Institute at Dublin City University have developed an inexpensive printable UV dosimeter for SODIS applications that can be read using a **mobile phone**.

Context Words: - *smartphone, ipad, cellular, devices, sms, tablets, laptops, landline, cloud, portable, blackberry, nokia, phone, personal, satellite, network, telephone, camera voice, telecommunication, touch, screen, touchscreen, printers. Wired, telecom, messaging, pocket etc.*



Threshold plot for cluster 0

Cluster 1

Implied Meaning : - “Portable”

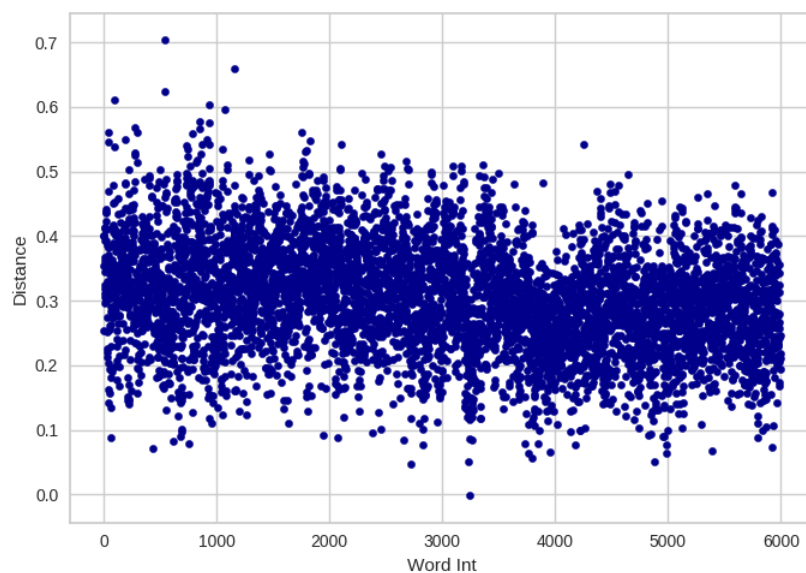
Example Sentences: -

- 1) One way of attributing methane to a particular source is taking continuous measurements of the stable carbon isotopic measurements of atmospheric methane $\delta^{13}\text{CH}_4$ in the plume of anthropogenic methane sources using a **mobile analytical system**.
- 2) For example, transcontinental gas pipelines, linked with regional collection and distribution networks, now spread throughout much of the world. Flare Gas Recovery Systems FGRS for processing APG into liquid or compressed fuels at the wellpad have also become **increasingly mobile** and varied in their capabilities.

3) Volume I has since included stationary point and area source emission factors, and Volume II includes **mobile source emission factors**.

4) In the case of mixed oxide MOX fuel, the xenon tends to diffuse out of the plutonium-rich areas of the fuel, and it is then trapped in the surrounding uranium dioxide. The neodymium tends to **not be mobile**.

Context Words: - flexible, portable, anchored, trapped, migration, diffuse, motion, permanent, clamped, location, stuck, suspended, soluble, dense, constrained, static, rotatory, coiled, wireless, wired, dispersion, remote, transverse, floating etc.



Threshold plot for cluster 1

Target word: - plant

No of clusters: - 2

Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

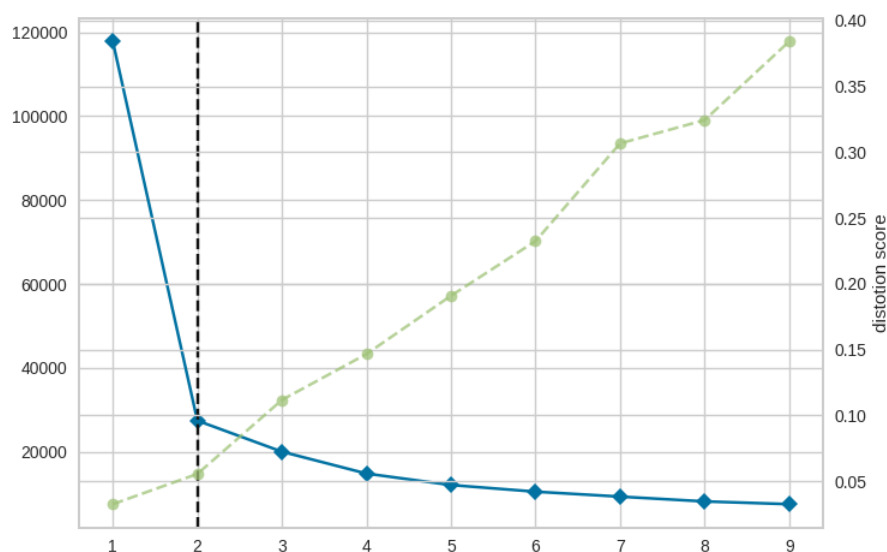
YES

Predicted Clusters by the elbow method = $k = 2$

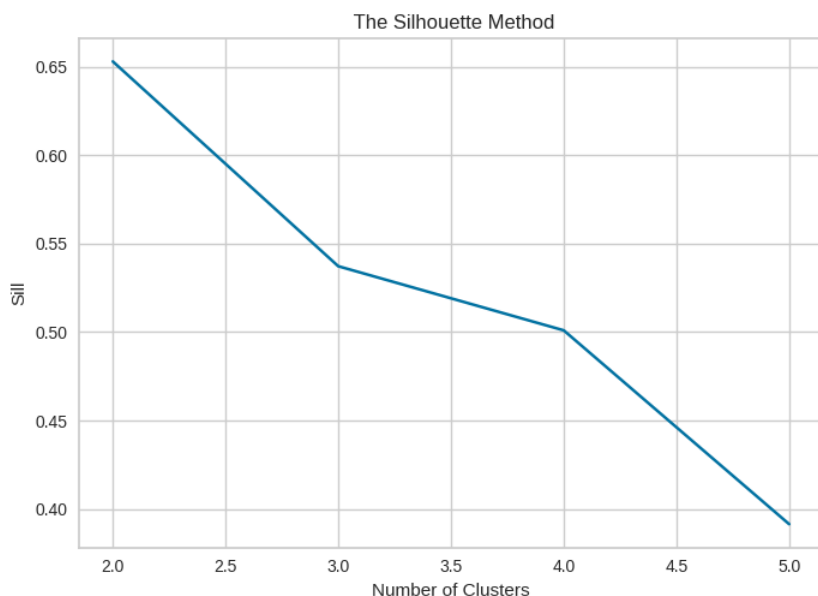
Predicted clusters by the silhouette method = 2

Sil Score for $K = 2 = 0.65$

WCSS for the elbow method for $K = 2$ is 3000 and distortion score is 0.010



Elbow Plot



Silhouette Plot

Cluster 0

Implied Meaning: - “(botany) a living organism lacking the power of locomotion”

Example Sentences: -

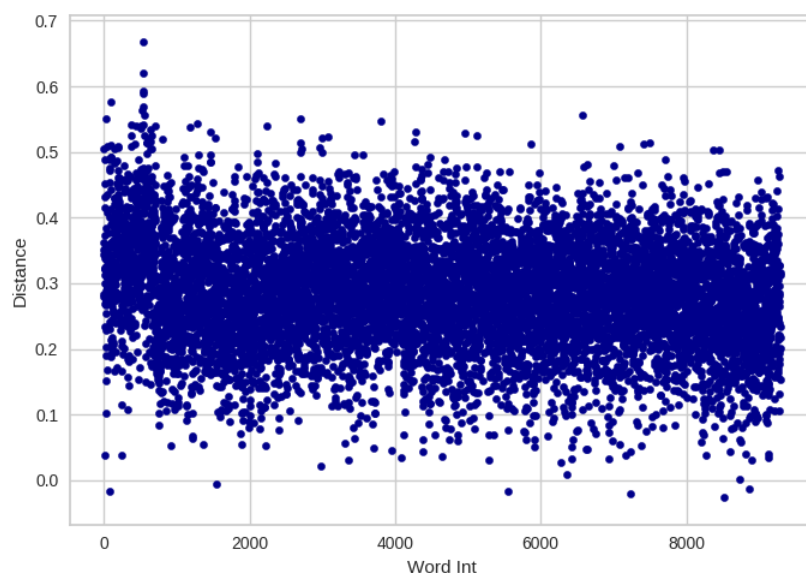
1) [34] Before invention discovery concept "nature" ancient Greek physis Pre-Socratic philosophers, words tend used describe natural "way" **plant grows**, [35] "way" which, example, one tribe worships particular god. For reason, claimed men first philosophers strict sense, also first people clearly distinguish "nature" "convention."

2) It is sometimes called the central science because it provides a foundation for understanding both basic and applied scientific disciplines at a fundamental level.[4] For example, chemistry explains aspects of **plant chemistry** (botany), the formation of igneous rocks (geology), how atmospheric ozone is formed and how environmental pollutants are degraded (ecology), the properties of the soil on the moon (astrophysics), how medications work (pharmacology), and how to collect DNA evidence at a crime scene (forensics).

3) Berthouly-Salazar, Cécile; Vigouroux, Yves; Billot, Claire; Scarcelli, Nora; Jankowski, Frédérique; Kane, Ndjido Ardo; Barnaud, Adeline; Burgarella, Concetta (2019). "Adaptive Introgression: An Untapped Evolutionary Mechanism for Crop Adaptation". *Frontiers in Plant Science*.

4) In addition, biologist Matthew Parker, after reviewing numerous genetic studies on **plant disease** resistance, failed to find a single example consistent with the concept that pathogens are the primary selective agent responsible for sexual reproduction in the host.

Context Words:- agriculture, vegetation, botany, stems, animal, fungal, vine, vegetable, herbs, trees, weeds, soil, shrub, fungi, flowers, leaf, wild, woody, animals, bushes, industrial, fruits, trees, leaves, grassland, fish, biological, geographical, bacterial, environment, moisture, foilage etc.



Threshold Plot for cluster 0

Cluster 1

Implied Meaning : - “buildings for carrying on industrial labor”

Example Sentences: -

1) It is essentially a modified version of their nuclear propulsion system with heat generated by ethanol and oxygen. Specifically, a conventional steam turbine **power plant** is powered by steam generated from the combustion of ethanol and

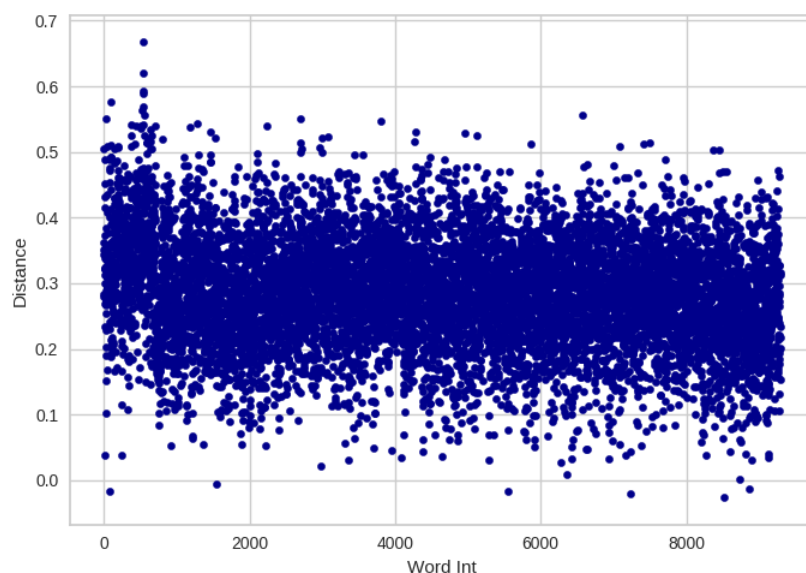
stored oxygen at a pressure of 60 atmospheres. This pressure-firing allows exhaust carbon dioxide to be expelled overboard at any depth without an exhaust compressor.

2) Such a propulsion system has a smaller footprint than a diesel-only **power plant** with the same maximal power output, since smaller engines can be used and the gas turbine and gearbox don't need that much additional space.

3) Some nuclear submarines also use a similar turbo-electric propulsion system, with propulsion turbo generators driven by **reactor plant steam**. Among the pioneering users of true diesel-electric transmission was the Swedish Navy with its very first submarine, HMS Hajen later renamed Ub no 1, launched in 1904 and originally equipped with a semi-diesel engine a hot-bulb engine primarily meant to be fueled by kerosene, later replaced by a true diesel.

4) More recent prototypes include the SEP modular armoured vehicle and T95e. Future tanks may use diesel-electric drives to improve fuel efficiency while reducing the size, weight and noise of the **power plant**.

Context Words:- facility, stations, generating, chamber, equipment, columns, facilities, stack, terminal, station, usage, boiler, vessel, ta, generator, rod, power, loop, factory, industry, reforming, reactor, turbines, rack, stripes, pipeline, machinery, installed, line, depot, grid etc.



Target word: - Drive

No of clusters: - 4

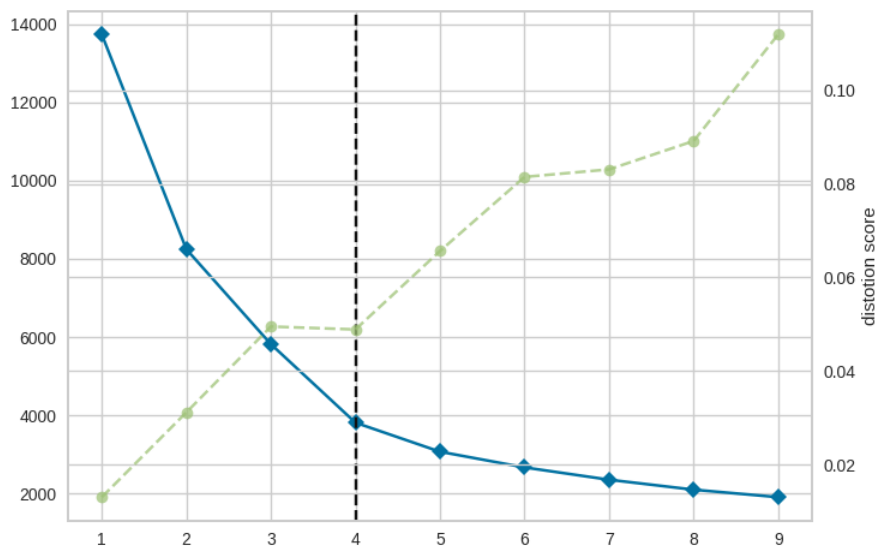
Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

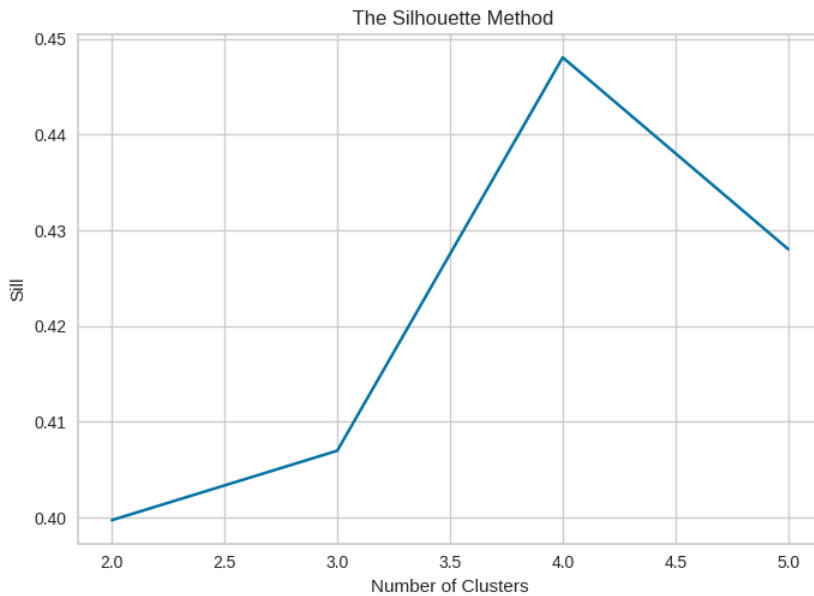
Do the number of clusters make sense?

YES

k = 4



Elbow Plot



Silhouette Plot

Cluster 0

Implied Meaning: - “strive and make an effort to reach a goal ”

Example Sentences: -

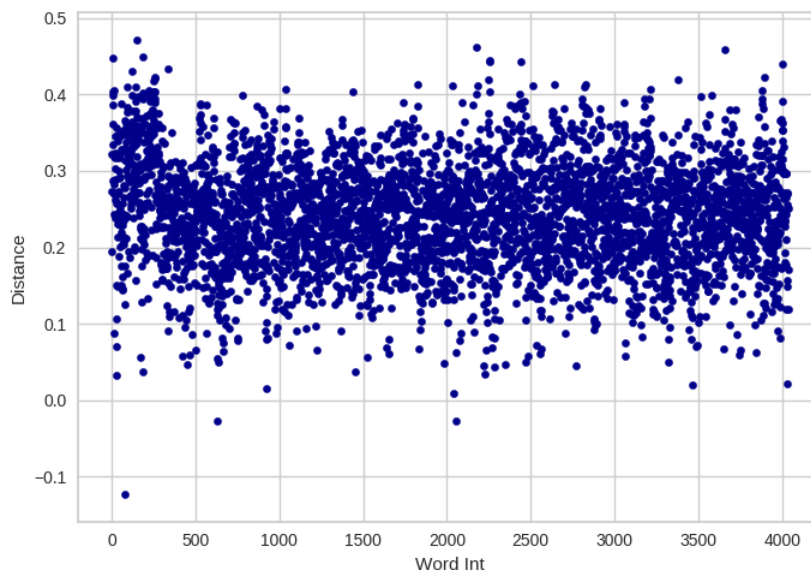
1) Analog Behavioral Modeling with the Verilog-A Language by Dan FitzPatrick, Ira Miller. Mark Bew MBE born 5 March 1967 is an English engineer and chairman of the PCSG consultancy business, formerly ECS. Until January 2015, he was also chairman of BuildingSMART's UK chapter, and, from 2011 to 2016, was chair of the UK Government's BIM Task Group, a body created to drive implementation of building information modelling BIM across UK public sector construction projects.

2) Sensors and instrumentation **drive the central forces** of innovation, not only for Industry 4.0, but also for other “smart ” megatrends, such as smart production, smart mobility, smart homes, smart cities and smart factories.

3) Even though CAE has built a strong reputation as a verification, troubleshooting and analysis tool, there is still a perception that sufficiently accurate results come rather late in the design cycle to really **drive the design**.

4) This is combined with intelligent reporting and data analytics. The objective is to let simulation **drive the design**, to predict product behavior rather than to react on issues which may arise, and to install a process that lets design continue after product delivery.

Context Words: - driving, driven, swing, push, adjusting, hit, pumped, run, cruise, impetus, track, predict, deploy, pitch, manage, react, lead, persist, push, encourage, decide, drill, retrieve, generate, develop, pushing, stimulate, pilot, enables, triggered, continue, governing etc.



Threshold Plot for cluster 0

Cluster 1

Implied Meaning: - “cause to function by supplying the force or power for or by controlling”

Example Sentences: -

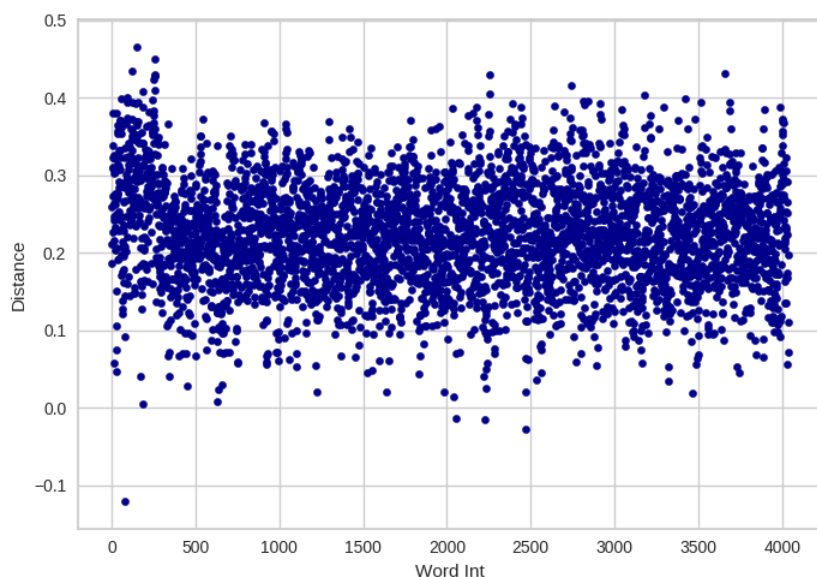
1) Unlike the azimuth thruster where a conventional propeller is rotated about the vertical axis to direct its thrust, allowing a vessel to steer without the use of a rudder, the Voith-Schneider **drive merely** requires changing the pattern of orientation of the vertical blades.

2) This form of power transmission is called a Z-drive because the rotary motion has to make two right angle turns, thus resembling the letter "Z". This name is used to differentiate the arrangement of drive to that of the **L-drive**.

3) The advent of the **stern drive** propulsion leg improved design so that auto engines could easily power boats. Large craft For larger craft, including ships where outboard propulsion would in any case not be suitable the propulsion system may include many types, such as diesel, gas turbine, or even fossil-fuel or nuclear-generated steam.

4) In 1894, Hungarian engineer Kálmán Kandó developed a new type 3-phase asynchronous electric **drive motors** and generators for electric locomotives.

Context Words: - motor, axle, gear, gearbox, fan, valve, reversal, thrust, feed, transmission, steamer, mechanism, shaft, bracket, springs, screw, dam, mounted, grid, assembly, propulsion, helix, cylindrical, steering, clip, geared, operated, belt etc.



Threshold plot for cluster 1

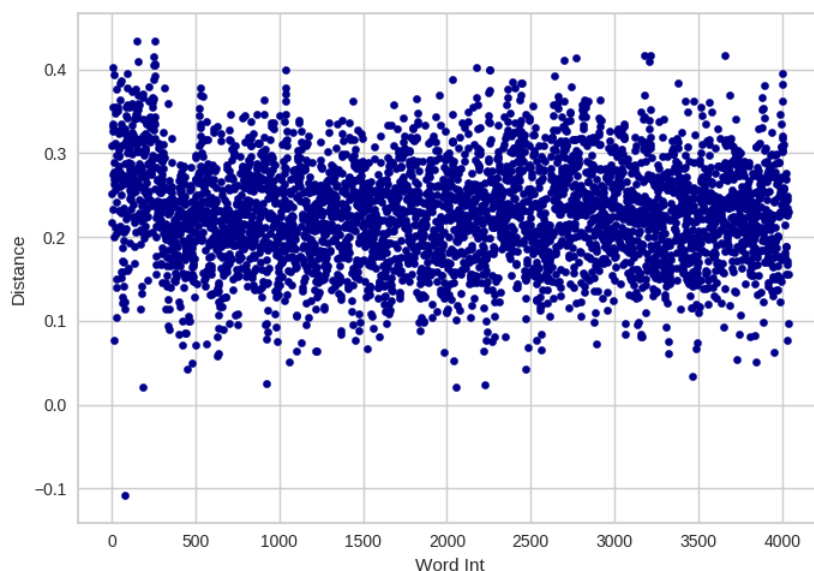
Cluster 2

Implied Meaning: - “To drive a vehicle”

Example sentences:-

- 1) Named after German aviation engineer and inventor Anton Flettner: In a rotor ship the rotors stand vertically and lift is generated at right angles to the wind, **to drive** the ship forwards.
- 2) Phosphenes are spread out across the visual field in what researchers call the starry-night effect. Immediately after his implant, Naumann was able to use his imperfectly restored vision to **drive slowly** around the parking area of the research institute.
- 3) On the night of Helen's murder, the Kimbles had been heard, earlier the same day, arguing heatedly over this topic by their neighbors. Richard later went out for **a drive to cool off**; as he was returning home, he nearly struck with his car a man with only one arm, who was fleeing from the vicinity of the Kimble house. Richard then found that Helen had been killed.
- 4) But no one had seen or heard Richard go out for **his drive**, or seen him while he was out, and there was no evidence that the "one-armed man" Kimble saw existed.

Context Words:- driving, cars, highway, travelling, road, automobile, vehicle, travel, around, transit, road, headlights, racing, pull, cruise, parking, slowly, highways, roads, traffic, pavement, green, signal, destination, wash, station, carrier etc.



Threshold Plot for cluster 2

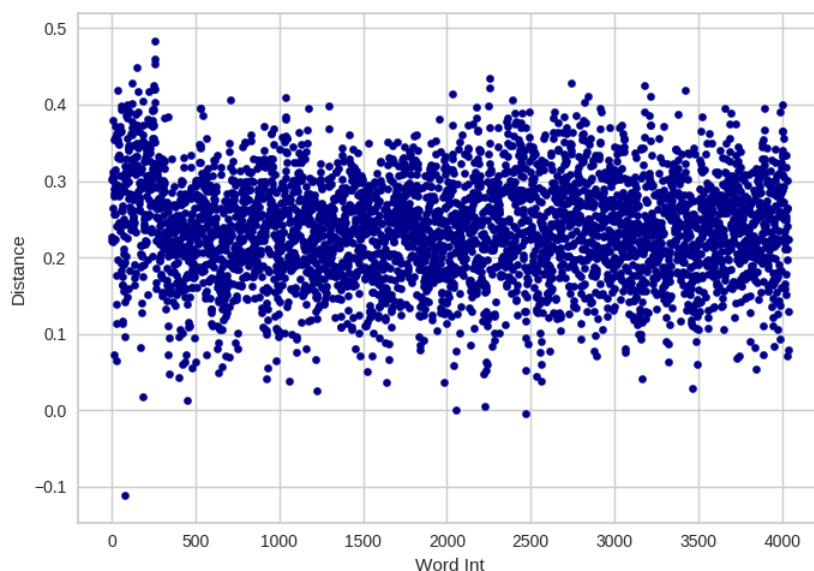
Cluster 3

Implied Meaning: - “A hard drive”

Example Sentences: -

- 1) In tape storage field they are known as tape libraries, and in optical storage field optical jukeboxes, or optical disk libraries per analogy. The smallest forms of either technology containing just one drive device are referred to as autoloaders or autochangers.*
- 2) Comparison of several forms of disk storage showing tracks (not-to-scale); green denotes start and red denotes end. Some CD-R(W) **Drive** and DVD-R(W)/DVD+R(W) recorders operate in ZCLV, CAA or CAV modes.*
- 3) The practitioners of big data analytics processes are generally hostile to slower shared storage,[52] preferring direct-attached storage (DAS) in its various forms from **solid state drive** (SSD) to high capacity SATA disk buried inside parallel processing nodes. The perception of shared storage architectures—storage area network (SAN) and network-attached storage (NAS)— is that they are relatively slow, complex, and expensive. These qualities are not consistent with big data analytics systems that thrive on system performance, commodity infrastructure, and low cost*
- 4) The disk drive industry has followed a different pattern. **Disk drive** capacity is generally specified with unit prefixes with decimal meaning, in accordance to SI practices.*

Context Words: - *partition, ram, discs, disk, memory, logical, optical, solid, state, cartridge, slot, cd, dvd, graphics, moving, hardware, file, manager, download, os, rom, kernel, processor, bus, readers, card, pen, chip, port, usb, slot, screen, transfer, communication, speed, format, stored, assembly, layout etc.*



Threshold plot for cluster 3

Target word: - Bank

No of clusters: - 4

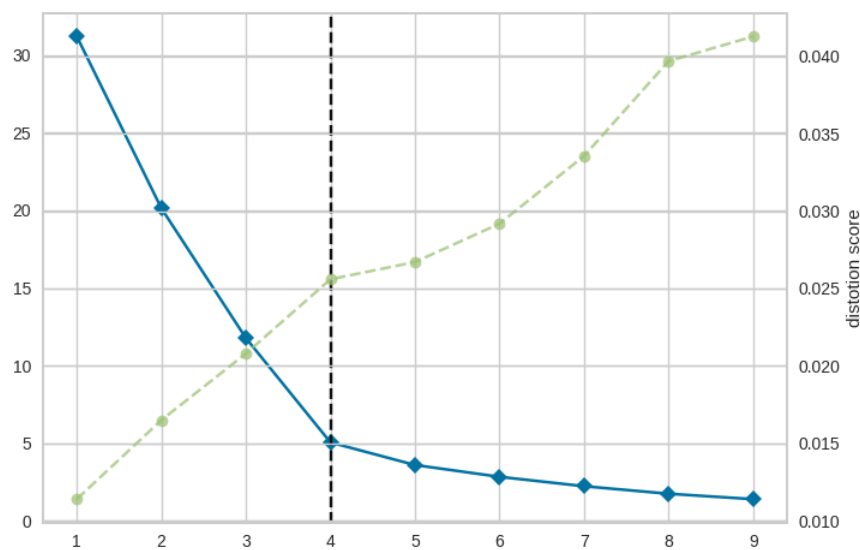
Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

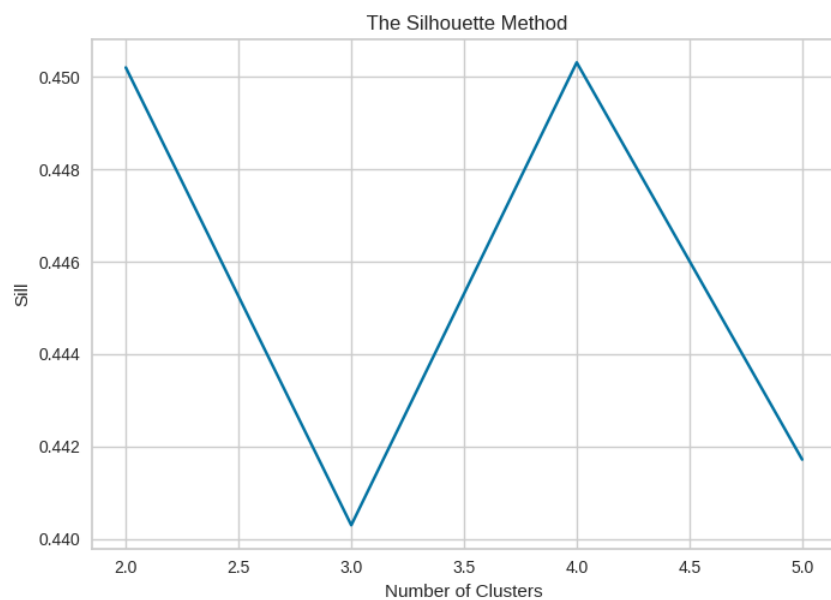
Do the number of clusters make sense?

YES

No of clusters in the CS Domain were 2 (Mainly in the context of a financial institution and a storage space). Extra two clusters have been observed in the mixed corpus.



Elbow Plot





Silhouette Plot

Cluster 0: -

Implied Meaning : - “Shore (Especially of a river)” (This is the extra meaning extracted from the corpus)

Example Sentences: -

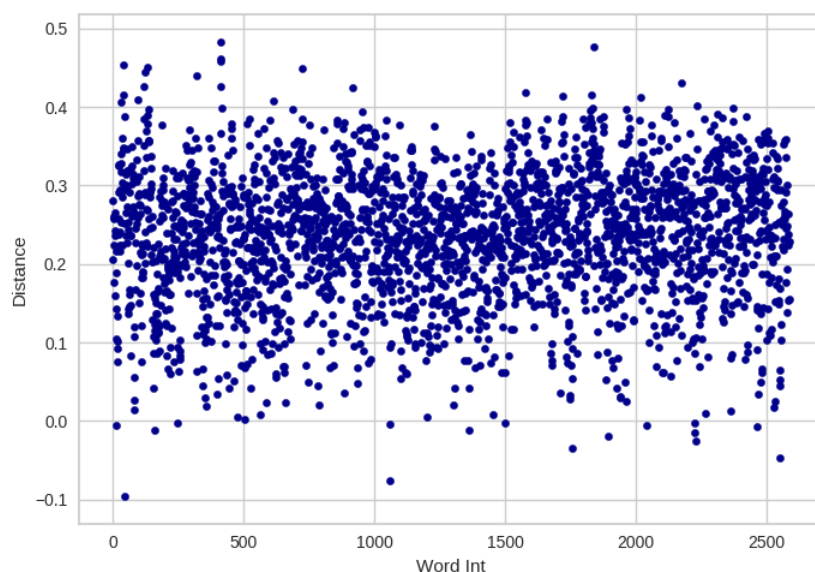
1) Location and geology. The Drake Well is located in Cherrytree Township, Venango County in northwestern Pennsylvania. situated on the flats 150 feet 46 m from the **east bank** of Oil Creek. The site was originally on an artificial island formed by the creek and a mill race.

2) It is not on the south side of Main Street, where a replica oil derrick stands. The site is in a small public park, about 1 acre 0.40 ha in size, set on the **western bank** of the Verdigris River. The park has open grassy areas and a small parking area, and the original site of the oil well is marked by a plaque.

3) Learning of Motonao's demise, he drew up his forces and marched to engage the smaller Mōri resistance. The Takeda encountered the Mōri and Kikkawa occupying the opposite **bank of the Uchikawa River** 又打川 and a bitter struggle ensued. Heavily outnumbered, the Mōri-led forces began to falter and fall back, but they held in place only by Motonari's pleas to stand their ground.

4) The launching arm features a large bar that features as an outrigger during emplacement. After assault force vehicles have passed over the bridge, the AVLB can cross over, pick up the bridge on **the far bank** and continue along in support of assault forces. It can produce a limited smokescreen by dumping raw diesel fuel into the exhaust system to visually obscure the area around the vehicle.

Context Words: - shore, bluff, ravine, downstream, faces, sides, confluence, hillside, delta, rainfall, reservoir, pond, rivers, channel, beach, downstream, flank, mouth, southern etc.



Threshold scatter plot for a label in cluster 0.

Cluster 1

Implied Meaning: - “a steep slope (as of a hill) or tilted inward slope” (Another unexpected context that has caused quite a stir in the expectations)

Example Sentence: -

1) The lucrative right to charge tolls Zollrecht, often in connection with the right to sell alcohol Krugrecht could be given to local farmers. Construction

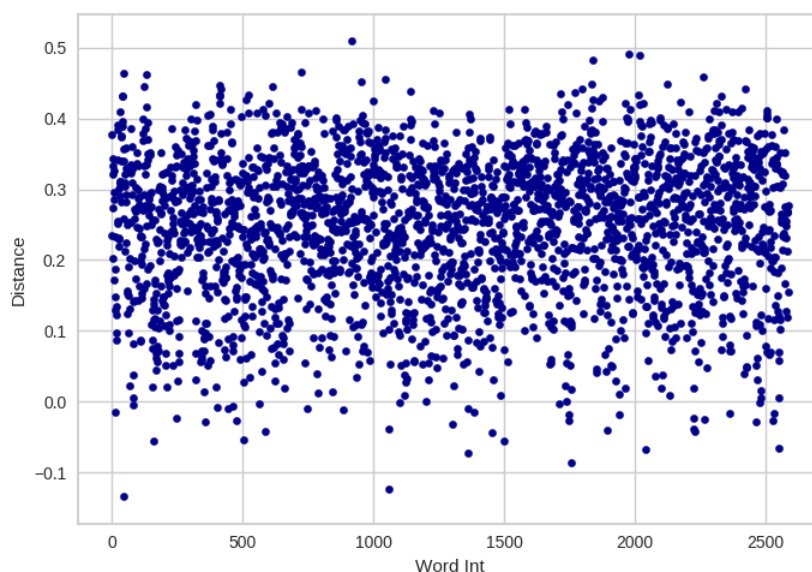
Most landwehrs consisted of a simple ditch as an obstacle - on the plains this was also a water channel or moat - behind which was an earthen **bank made** of the spoil from excavating the ditch. **Behind the bank** was the main barrier itself, a 20 to 50 metre-wide, dense, impenetrable thicket.

2) **Behind the bank** was the main barrier itself, a 20 to 50 metre-wide, dense, impenetrable thicket. In upland areas, the route of the landwehr conformed to natural features such as rocks, steep slopes and watercourses.

3) In particular, double ditches were designed to prevent them from being jumped by horsemen. Another design was the so-called Wehrhecke Wallhecke or Knick, **a hedge-planted bank**, for the maintenance of which a tax called Knickgeld was raised.

4) In Hesse, at the end of the 17th century, many villages on important roads or by borders, had fortifications, excluding any fortified churches, as a Hessian chronicler described in 1697: Today is almost the majority of large districts and villages in Hesse are surrounded by a moat and embankment / so that they can defend themselves against minor attacks. The fortification could be made of fences known as Etter or Dorfetter, hedges, **bank and ditch** dry as well as wet and gates. Another temporary artificial obstacle which could be built relatively quickly in the event of an attack, or to close gaps in a landwehr, was the abatis.

Context Words: - channel, bed, loft, rise, anflod, depression, base, fencing, moat, dyke, barrier, shaft, slip, strip, territory, spiral, holes, guts, walls, stones etc.



Threshold Cluster for a label in cluster 1

Cluster 2



Implied Meaning: - “Financial and a monetary institution”

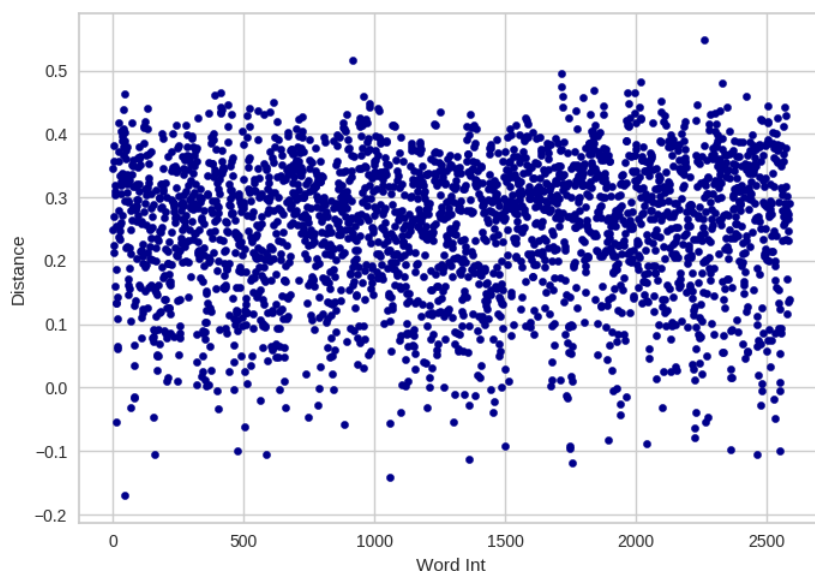
Example Sentences: -

- 1) Marcos Barbosa Pinto was the chief of staff for Demian Fiocca, President of BNDES during the Luis Inácio Lula da Silva administration. He was invited to serve as **president of the bank**
- 2) The MBRT evaluates and permits a proposed Mitigation Bank. They also determine the number of potential mitigation **credits a bank** may earn and sell. The MBRT may include representatives of various federal, state and/or local government agencies, including: U.S. Army

Corps of Engineers, National Marine Fisheries Service, Environmental Protection Agency, US Fish and Wildlife Service, State Environmental Protection Divisions, Local Water Management Districts, County Environmental Departments and the Soil Conservation Service.

- 3) **A mitigation bank** generates credits for the amount and quality of habitat the bank site improves. Credits are units of exchange defined as the ecological value associated with converting a naturally occurring wetland or other specific habitat type, for economic purposes.
- 4) The estimated potential number of credits a bank may earn can vary based on ecological **performance of the bank**. The Interagency Review Team periodically releases bank credits as the bank meets certain performance milestones. This happens over the course of the "establishment period", which lasts around 10 to 12 years.

Context Words : - banking, institution, financial, stakeholder, stocks, credit, market, sell, government, financing, firms, union, account, centers, reserves, money, monetary, payment, loans, cards, deposits etc.



Threshold Cluster Plot for a label in cluster 2

Cluster 3

Implied Meaning : - “A storage unit”

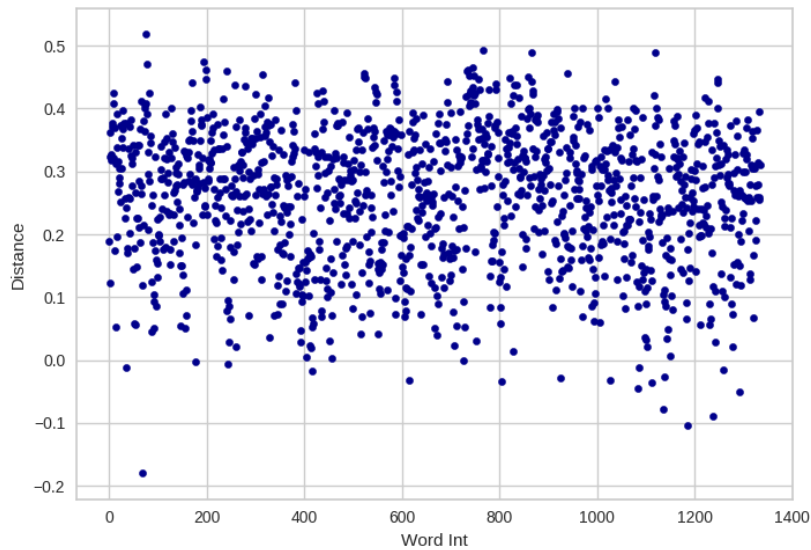
Example Sentences: -

1) Fuel cells or flow batteries may provide significant advantages in years to come. Today 2017 however they are still expensive and require specialist equipment and knowledge. The size of the **battery bank** determines the range of the boat under electric power. The speed at which the boat is motored also affects range – a lower speed can make a big difference to the energy required to move a hull.

2) **Battery bank**, There have been significant technical advances in battery technology in recent years, and more are to be expected in the future. Lead–acid batteries were still the most viable option until the advent of larger, lithium-ion batteries mass-produced for electric cars from approximately c2012 onwards.

3) This can be seen on this footage **video bank** effects taken during model tests in a towing tank at Flanders Hydraulics Research

This phenomenon depends on many parameters, such as bank shape, water depth, ship-bank distance, ship properties, ship speed and propeller action.



Threshold scatter plot for cluster 3

Section 2: - A Few More Examples

Target word: - Application

No of clusters: - 2

Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

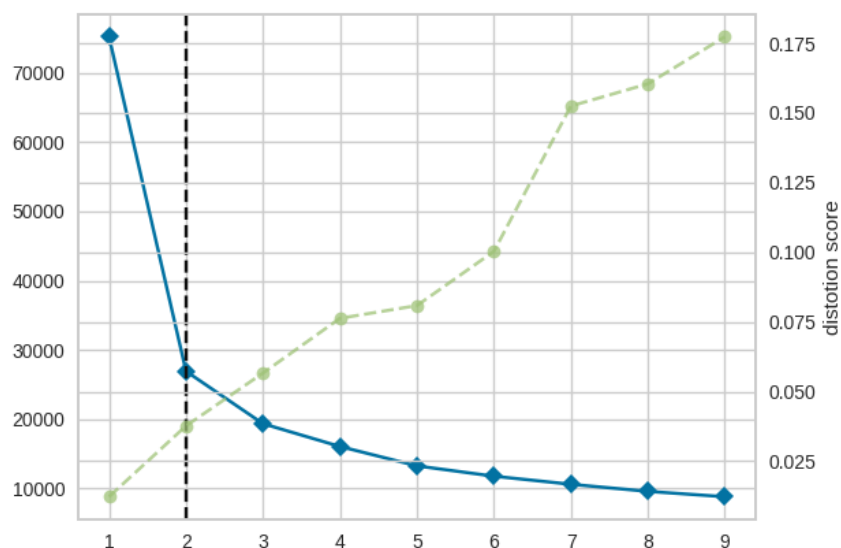
Do the number of clusters make sense?

YES

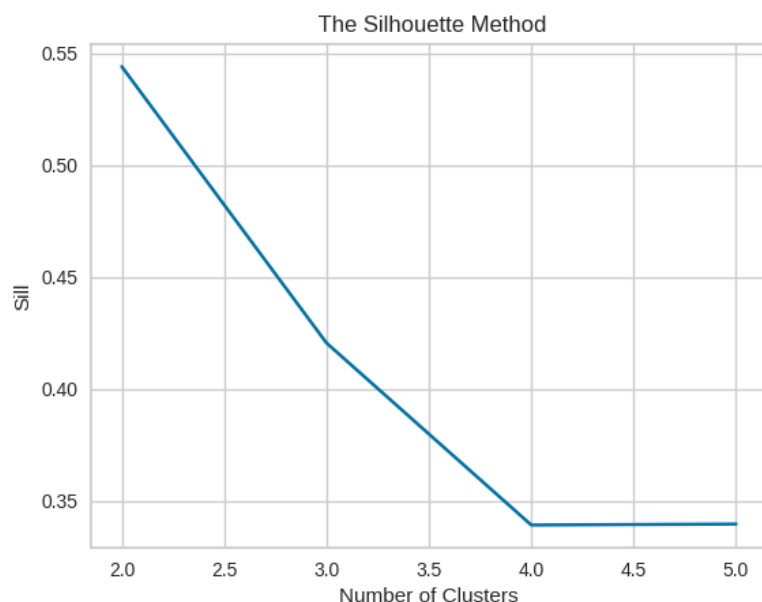
Elbow = Silhouette = 2

Silhouette Score for K= 2 is 0.54 (This is the maximum silhouette score that was obtained)

The observed WCSS score for the elbow method was found out to be 4000 for the mixed corpus. It was around 2500 in the CS domain. An outright comparison wouldn't be fair for the two scores, as the no samples increase drastically, and the scores obtained were at the knee point of the respective cluster graphs. The information is provided just to visually juxtapose the two cluster graphs.



Elbow Plot



Silhouette Score for the mixed corpus

Cluster 0: -

Implied Meaning: - “Apply”, “Implement”

This cluster contains all the sentences in the corpus that use the word in the context to apply. The predicted context words are of a decent quality. Not extremely outstanding, but reading a few of them presents a general idea of the context

Example Sentences :-

1) Many modern advancements in technology have been inspired by nature and natural systems, including aircraft, automobiles, architecture, and even industrial systems. Advancements in nanotechnology initiated the **application** of this technique to micro- and nano-scale problems, including tissue engineering. This technique has been used to develop synthetic bone tissues, vascular technologies, scaffolding materials and integration techniques, and functionalized nanoparticles.

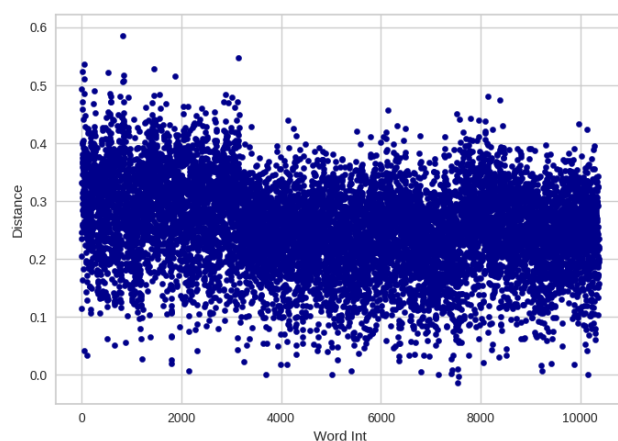
2) He has two doctoral degrees, two post-doctoral degrees, and has authored more than 300 research articles and 20 books/book chapters. Prof. Ricardo Armentano has done a thorough and remarkable work in the field of Engineering in

Biology and Medicine from its theoretical fundamentals to its **application** in clinical practice, throughout a conscious-technology approach with humanistic motivation and global vision.

3) His research investigates the **application** of engineering principles to control tissue assembly, repair and regeneration, and incorporates areas including nanotechnology, tissue engineering, engineered cellular micro-environments, micro-electromechanical systems and micro-fabrication technologies. Chen has been awarded numerous awards and distinctions such as ONR Young Investigator Award in 1999, the Presidential Early Career Award for Scientists and Engineers PECASE in 2000,

4) Using CRISPR/Cas9 gene-editing, they mutated the gene called CCR5, which would confer resistance to HIV infection. The People's Daily announced the result as "a historical breakthrough in the **application** of gene editing technology for disease prevention"

Context words: - utilization, implementation, effectiveness, feasibility, incorporation, adoption, applying, investigation, integration, adapt, demonstrations etc.



Cluster Threshold plot for cluster 0

Cluster 1

Implied Meaning :- “App”, “Software Program”

The context meaning is same as that of cluster 0 of CS domain. There is no other context of the word application apart from these two in the mixed domain.

Example Sentences: -

1) It is showed in practical applications, that it was too complicated to handle in the field. In 1994 Siemens proposed a modified **application layer** with the name Decentralized Periphery DP which reached a good acceptance in the manufacturing industry. 2016 the Profibus is one of the most installed fieldbuses in the world and reaches 60 millions of installed nodes in 2018.

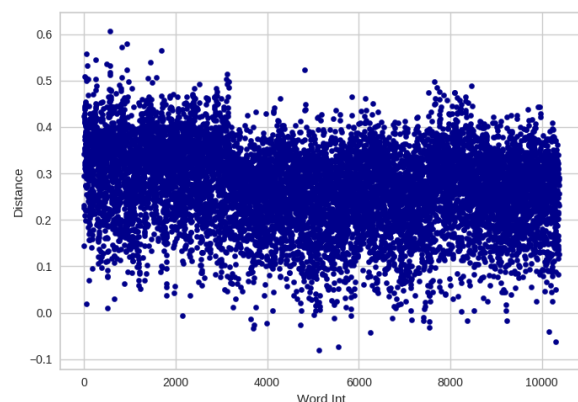
2) Either database **payroll application** also initiate process exchanging data ship shore, data transfer task often contain many components. These layers produce effect isolating implementation details one component assorted internal methods others.

3).If one requires differentiated hierarchy animals – differentiate, say, provide milk provide nothing except meat end lives – intermediary level abstraction, probably Dairy Animal cows, goats would eat foods suitable giving good milk, MeatAnimal pigs, steers would eat foods give best meat-quality. Such abstraction could remove need **application coder** specify type food, s/he could concentrate instead feeding schedule. The two classes could related using inheritance stand alone, programmer could define varying degrees polymorphism two types.

4) A **commercial application** that can be purchased from the iTunes App Store for displaying the entire BIMx Hyper-model: not only the 3D model of the building but also its two dimensional architectural documentation set as drawn and detailed with ARCHICAD.

Context Words: - gui, software, tool, deployment, apps, designer, programmer, program, widget, users, drivers, developer, configuration, installed, framework, daemon, rpg etc.

The words that have appeared are of a rich quality. Not only it has captured the sentences used in this context from the cs domain, but also from other domains as well. This gives an idea regarding the flexibility of the model.



Threshold Cluster plot for the word application

Target word: - Function

No of clusters: - 3

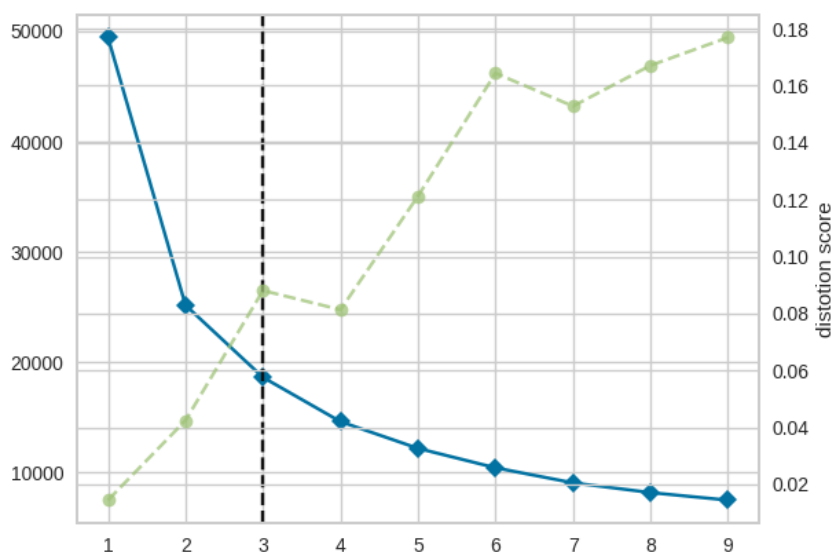
Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

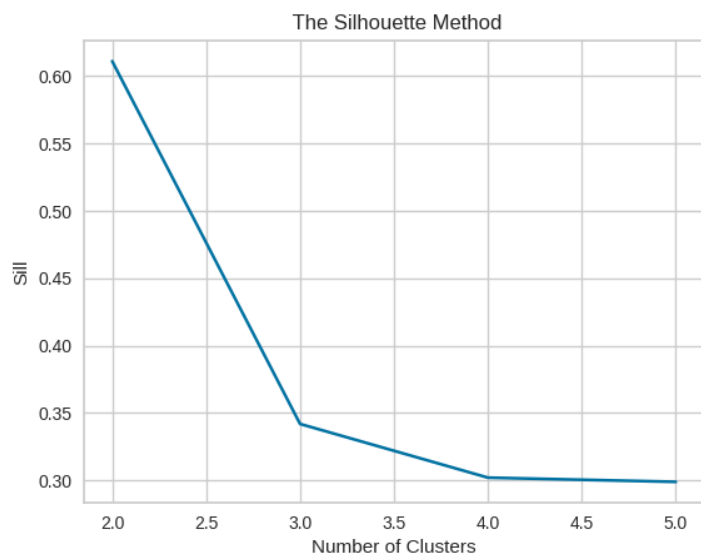
Do the number of clusters make sense?

YES

The Contexts found in the mixed corpus are similar to the contexts found in the CS domain. The word function has been used in 3 two contexts. Typically a mathematical or a programming function and the second as a verb in context to operate.



Elbow Graph



Silhouette Graph

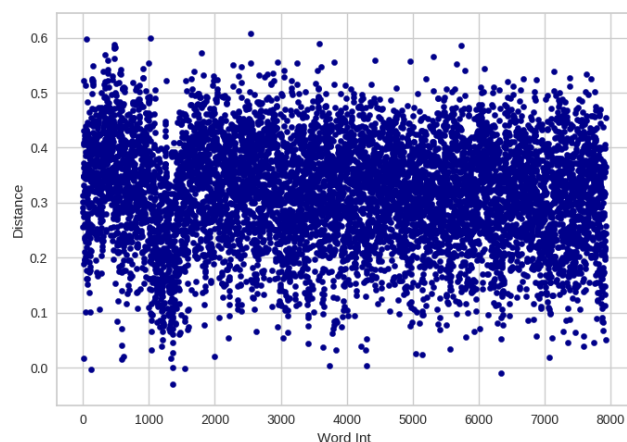
Cluster 0 :-

Implied Meaning:- “Mathematical Function”

Example Sentences: -

- 1) **Functions** were originally the idealization of how a varying quantity depends on another quantity. For example, the position of a **planet** is a function of time
- 2) If the **function** is called f , this relation is denoted by $y = f(x)$ (which reads "f of x"), where the element x is the argument or input of the function, and y is the value of the function, the output, or the image of x by f . [2] The symbol that is used for representing the input is the variable of the function (e.g., f is a function of the variable x). [3]
- 3) For example, if f is a **function** that has the real numbers as domain and codomain, then a function mapping the value x to the value $g(x) = 1/f(x)$ is a function g from the reals to the reals, whose domain is the set of the reals x , such that $f(x) \neq 0$.
- 4) Some widely-used **functions are** represented by a symbol consisting of several letters (usually two or three, generally an abbreviation of their name). In which case, a **roman type** is customarily used instead, such as "sin" for the **sine function**, in contrast to italic font for single-letter symbols.

Context Words: - curve, variable, map, range, domain, exponent, graph, line, linear, square, mapping, area, prediction, model etc.



Threshold scatter cluster plot for the cluster 0

Cluster 1 :-

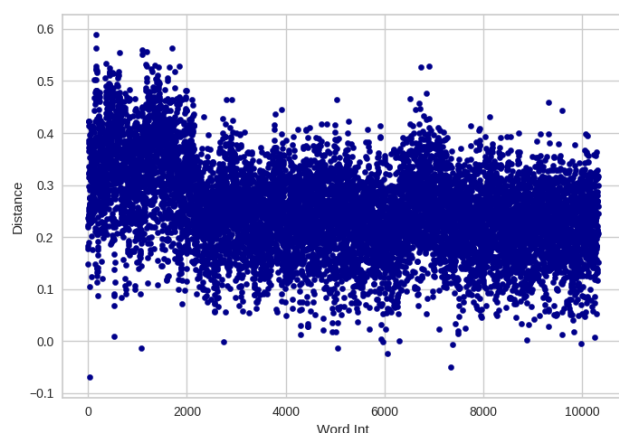
Implied Meaning:- “Programming Function”

Example Sentences: -

- 1) A typical use of a **function object** is in writing [callback](#) functions. A callback in [procedural languages](#), such as [C](#), may be performed by using [function pointers](#).
- 2) Another advantage of function objects is their ability to maintain a state that affects operator() between calls. For example, the following code defines a generator counting from 10 upwards and is invoked 11 times.
- 3) Java has no first-class functions, so function objects are usually expressed by an interface with a single method (most commonly the Callable interface), typically with the implementation being an anonymous inner class, or, starting in Java 8, a lambda.
- 4) In Lisp family languages such as Common Lisp, Scheme, and others, functions are objects, just like strings, vectors, lists, and numbers. A closure-constructing operator creates a function object from a part of the program: the part of code

given as an argument to the operator is part of the function, and so is the lexical environment: the bindings of the lexically visible variables are captured and stored in the function object, which is more commonly called a closure

Context Words: - callback, variable, method, object, java, class, argument, interpreter, code, programming, c, call, pass, loop, compiler, compile, model, input etc.



Threshold cluster plot for the cluster 1

Cluster 2: -

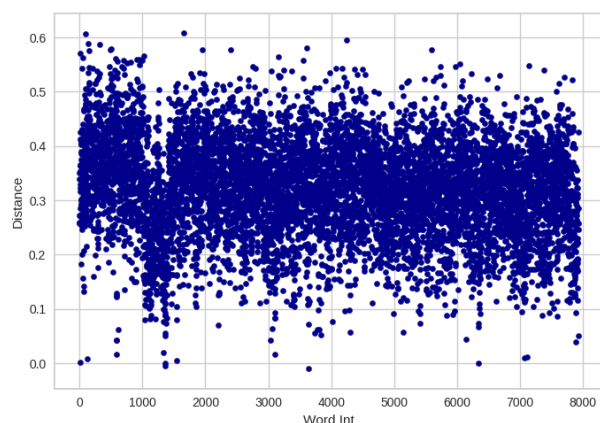
Implied Meaning: - “Operate”

Example Sentences: -

- 1) This allows us link behavior **brain function** help understand information processed. Different types imaging techniques vary temporal time-based spatial location-based resolution.
- 2) Computational neuroscience study **brain function** terms information processing properties structures make nervous system. It subset field neuroscience, looks analyze brain data create practical applications.
- 3) The **primary function** router forward packet toward destination network, destination IP address packet. To this, router needs search routing information stored routing table.
- 4) Stable synapses forget less easily, also harder consolidate. One recent computational hypothesis involves cascades plasticity allow synapses **function** multiple time scales. [33] Stereochemically detailed models acetylcholine receptor-based synapse Monte Carlo method, working time scale microseconds, built.

Context Words: - Implement, perform, maneuver, operate, runs, ability, assist, proliferation, regulation, process, operation, implementation, interactions, behavior, driven, flow etc.

One can definitely say after reviewing the cluster files that the vocabulary of the predicted words has become extremely rich as compared to the CS Domain.



Threshold scatter cluster plot for cluster 2.

Target word: - Language

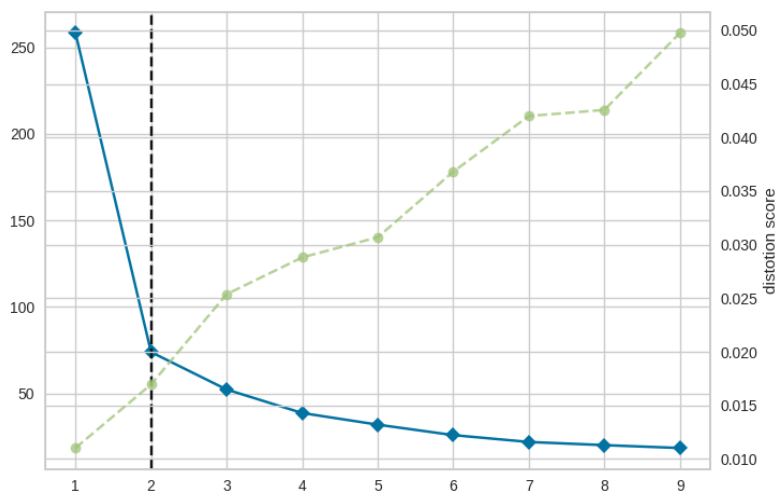
No of clusters: - 2

Deterministic Method :- C-Elbow Method

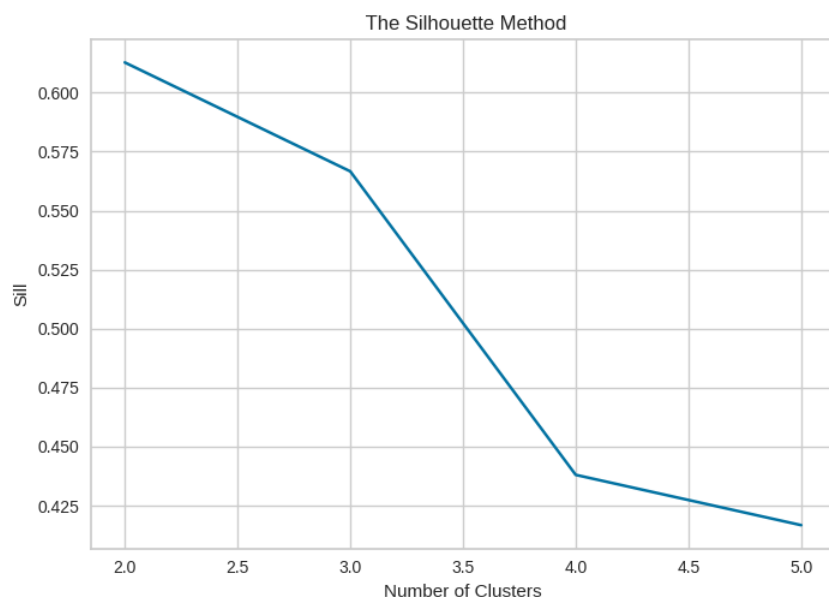
Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

YES



Elbow Plot



Silhouette Plot

Cluster 0

Implied meaning: - “Programming Language”

Example Sentences : -

1) To comply with these additional specifications, the implementation could use a linked list, or an array (with dynamic resizing) together with two integers (an item count and the array size). Functional-style ADT definitions are more appropriate for functional **programming languages**, and vice versa. However, one can provide a functional-style interface even in an imperative language like C. For example:

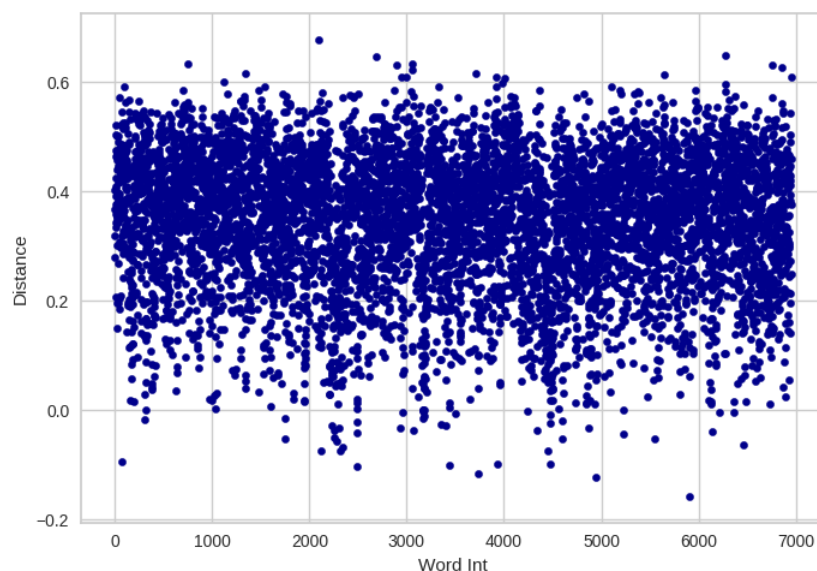
2) Compilers **language** processor is used for transforming source programs. An an interpreter computer software transforms and executes the indicated operations.

3) As computer technology provided resources, compiler designs could align better compilation process. It usually productive programmer use **high-level language**, development high-level languages followed naturally capabilities offered digital computers. High-level languages formal languages strictly defined syntax semantics form high-level language architecture.

4) ADTs are a theoretical concept, in computer science, used in the design and analysis of algorithms, data structures, and software systems, and do not correspond to specific features of.

computer languages—mainstream computer languages do not directly support formally specified ADTs.

Context Words :- library, variable, c, java, linguistic, package, module, class, interpreter, compiler, variables, system. Structure, engine, dialect, api, function, void, code, programmers, program, software etc.



Threshold Scatter plot for a cluster 0

Cluster 1

Implied Meaning: - “A medium of communication”

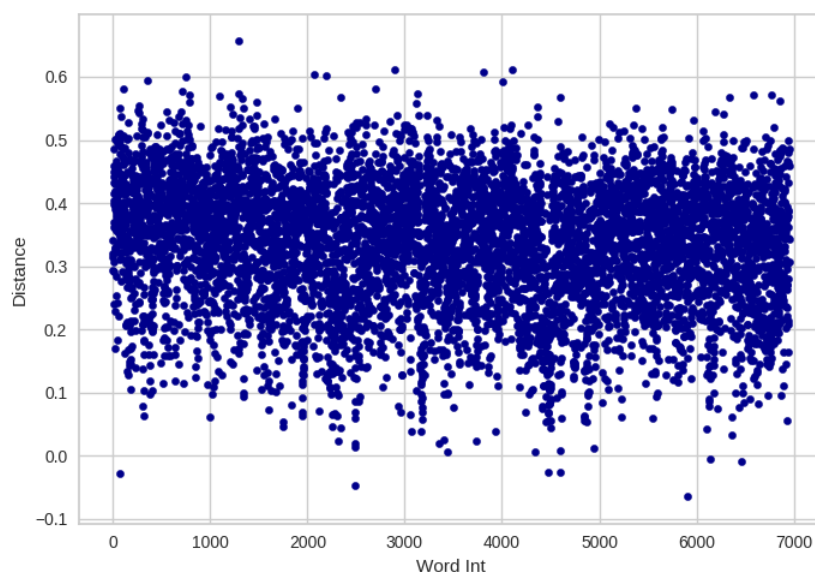
Example Sentences :-

- 1) The following list has commonly researched tasks like **natural language processing**. Some tasks are direct real-world applications, others commonly serve subtasks and are used to aid solving larger tasks.
- 2) The difficulty task depends greatly on complexity of morphology i.e., structure words in a 0 **language are considered**. English has a fairly simple morphology, especially inflectional morphology, thus it is often possible to ignore the task entirely and simply model the possible forms of word for e.g., "open, opens, opened, opening" separate words.

3) [dubious – discuss] Languages little inflectional morphology, English, particularly prone ambiguity. Chinese prone ambiguity **tonal language** verbalization. Such inflection readily conveyed via entities employed within orthography convey intended meaning. Parsing Determine parse tree grammatical analysis given sentence.

4) e.g., "close" root "closed", "closing", "close", "closer" etc.. Word segmentation Separate chunk continuous text separate words. **For language** like English, fairly trivial, since words usually are separated by spaces. However, written languages like Chinese, Japanese Thai mark the word boundaries by fashion, for languages text segmentation is a significant task requiring knowledge and vocabulary of morphology words in a language.

Context Words :- linguistic, sentence, grammar, grammatical, speaker, ling, consonant, vowel, spoken, comprehension, dialogue, pronounced, semantic, speech, facial, lexicon, dialect, understand, narrative etc.



Threshold Scatter Plot for a label in Cluster 1

Target word: - Float

No of clusters: - 3

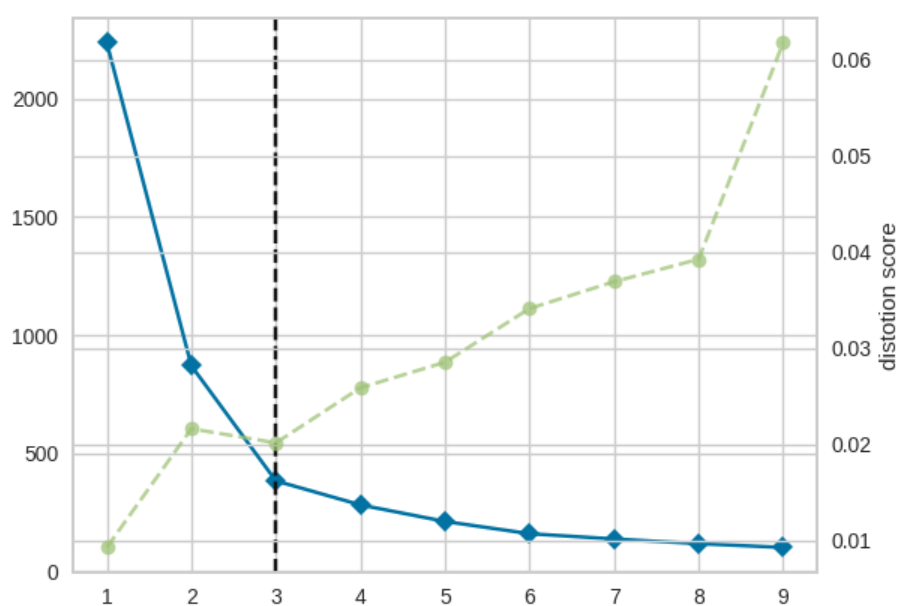
Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

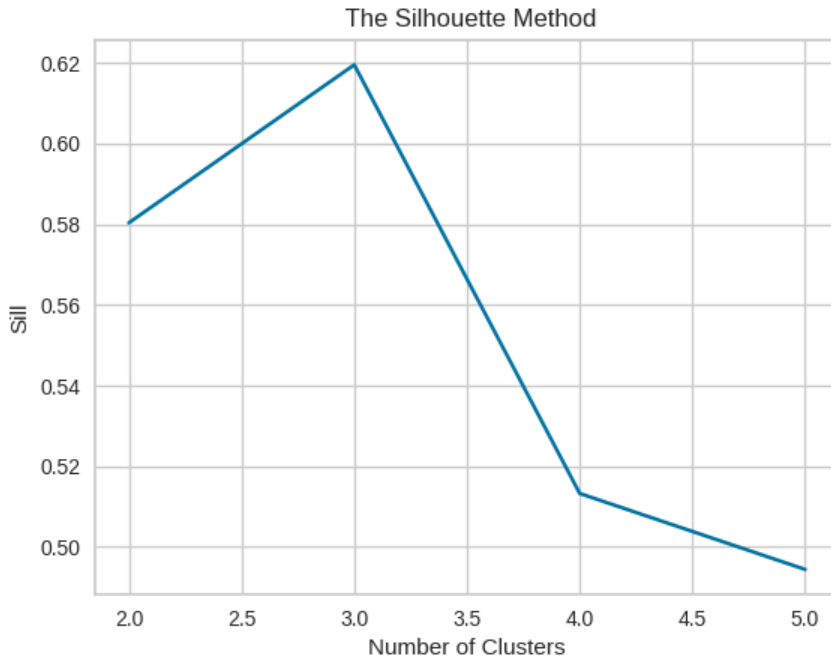
Do the number of clusters make sense?

YES

Clusters Predicted (K) = 3



Elbow Plot



Silhouette Plot

Cluster 0

Implied Meaning:- “An object that floats” (n)

Example Sentences: -

1) This bump can be seen at surface as a pressure spike at the cement pump. To prevent the cement from flowing back into the inside of the casing, a **float collar** above the casing shoe acts as a check valve and prevents fluid from flowing up through the shoe from the annulus.

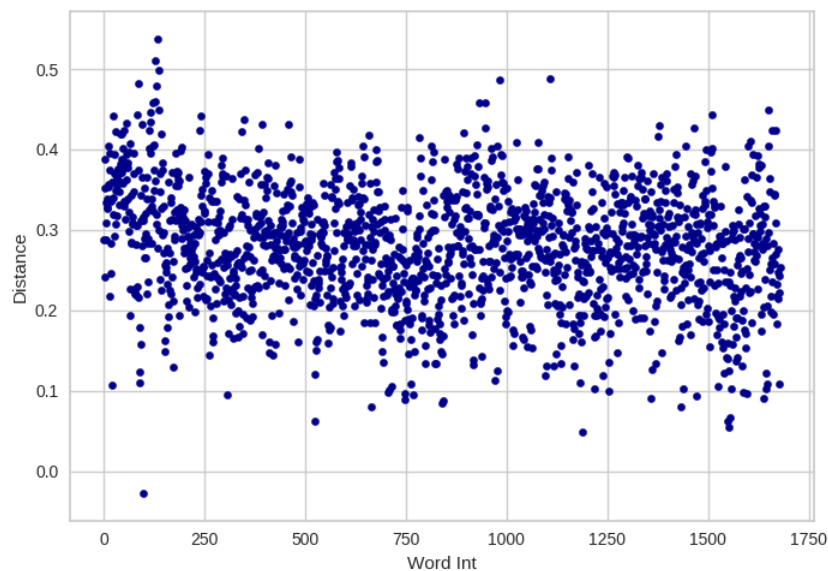
2) 35th Battalion, Singapore Combat Engineers 35SCE

35SCE is the battalion specialised in military bridging. Consisting of three companies, the battalion provides the transportation means in the form of **float bridges**, rafts and assault boats for the projection of combat troops and vehicles across rivers and water obstacles to facilitate troop movement.

3) Each bridging platoon operates 4 M3G **Float Bridges** which each require an operating crew of 4 men 2 sergeants and 2 Pioneer military. Each raft consists of two rigs, which form a Class 60 raft or a float bridge when coupled together further. The M3G Float Bridges also require substantial bank preparation prior to breaching in.

4) The sensor locked the cable reel so the falling anchor would pull the buoyant mine below the surface; and **the float extended** the antenna above the mine. Each mine had two hydrostatic safety features intended to render the mine safe if it detached from its mooring cable and floated to the surface.

Context Words: - tube, attachment, sink, hull, bar, island, shaft, upper, boats, carrier, submerged, keel, boat, pilot, floated, raft, vessel, dispersed, ship, structure, anchor, paddle, moore, sink, bubbles, boarding, flat, swim, hydro etc.



Threshold plot for cluster 0

Cluster 1

Implied Meaning: - “The act of floating” (v)

Example Sentences: -

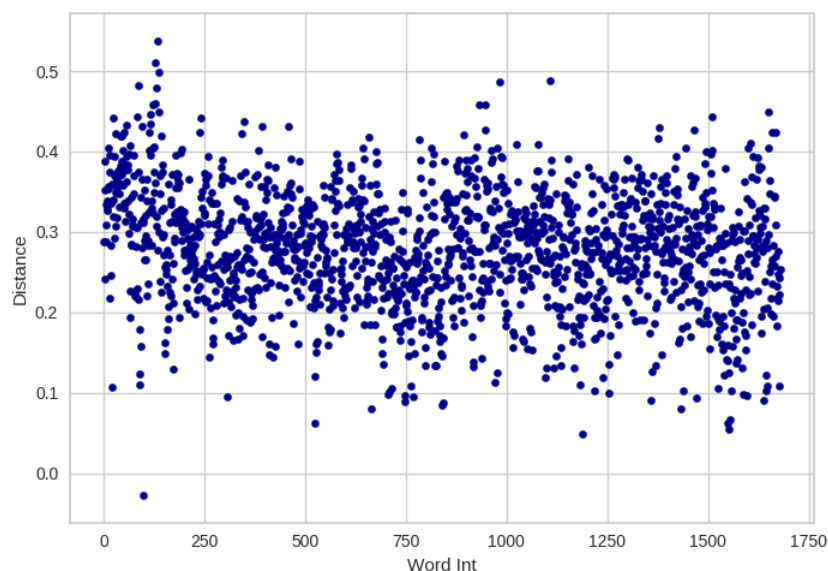
1) The up-flowing bubble design falls into three categories: those that attach to and shine through the aquarium or sump glass; those that **float on top** of the aquarium, sump or pond water surface; and those that go completely underwater like a submarine.

2) Petrol interceptors work on the premise that some hydrocarbons such as petroleum and **diesel float** on the top of water.

3) Significant in their "agitation froth flotation" process was the use of less than 1% oil and an agitation step that created small bubbles, which provided more surface to capture the **metal and float** into a froth at the surface. Useful work was done by Leslie Bradford at Port Pirie and by William Piper, Sir Herbert Gepp and Auguste de Bavay.

4) Frothing agents, known as frothers, may be introduced to the pulp to promote the formation of a stable froth on top of the flotation cell. The minerals that do **not float** into the froth are referred to as the flotation tailings or flotation tails.

Context Words: - floating, floated, bubble, aquarium, bottom, splash, drift, liquid, pond, swim, flowing, tank, anchored, settle, sail, submerged etc.



Threshold plot for cluster 1

Cluster 2

Implied Meaning: - "Data Type"

Example Sentences: -

1) The term floating point refers to the fact that a number's radix point (decimal point, or, more commonly in computers, binary point) can **"float"**; that is, it can be placed anywhere relative to the significant digits of the number. This position is

indicated as the exponent component, and thus the floating-point representation can be thought of as a kind of scientific notation..

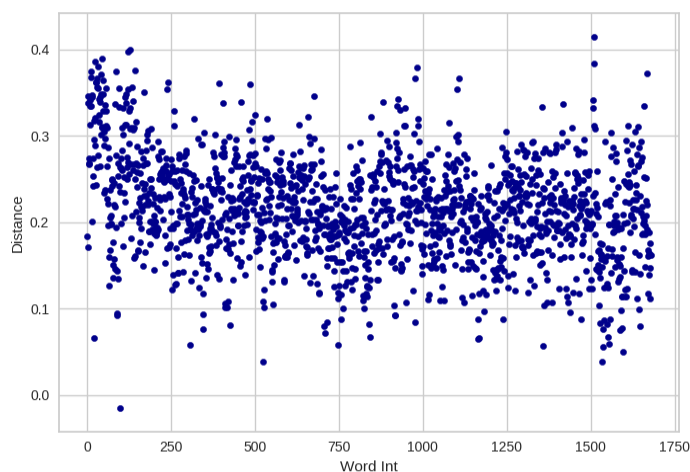
2) This example function in C to calculate the average of two numbers contains a logic error. It is missing parentheses in the calculation, so it compiles and runs but does not give the expected answer due to operator precedence (division is evaluated before addition).

```
float average(float a, float b)
{
    return a + b / 2;    /* should be (a + b) / 2 */
}
```

3) A compiler may use the static type of a value to optimize the storage it needs and the choice of algorithms for operations on the value. In many C compilers the float data type, for example, is represented in 32 bits, in accord with the IEEE specification for single-precision floating point numbers. They will thus use floating-point-specific microprocessor operations on those values (floating-point addition, multiplication, etc.)..

4) A union type definition will specify which of a number of permitted primitive types may be stored in its instances, e.g. "float or long integer". Contrast with a record, which could be defined to contain a float and an integer; whereas, in a union, there is only one type allowed at a time.

Context Words: - references, character, base, arguments, byte, operations, call, return, store, evaluated, pointer, strings, returns, element, primitive, integer, index, array, int, value, calculate, loop, syntax, save, char, null, name, decimal, etc



Threshold scatter plot for cluster 2

Target word: - Stack

No of clusters: - 2

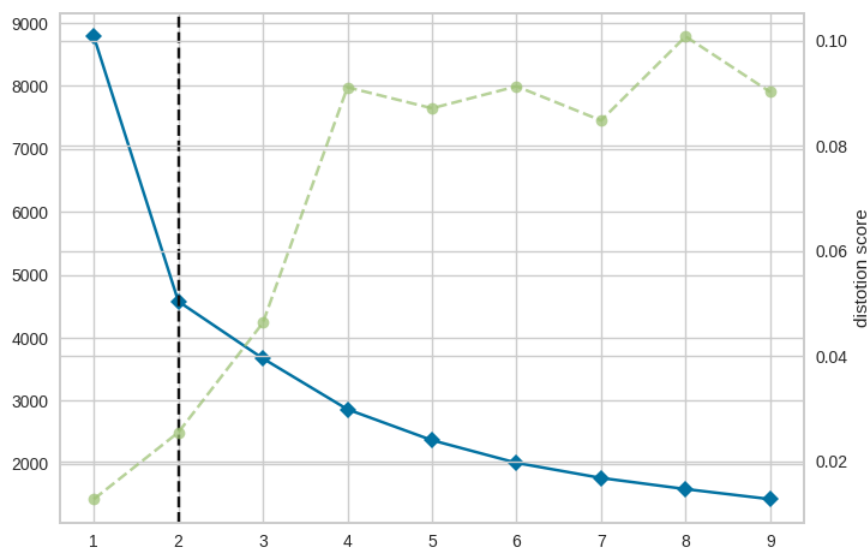
Deterministic Method :- C-Elbow Method

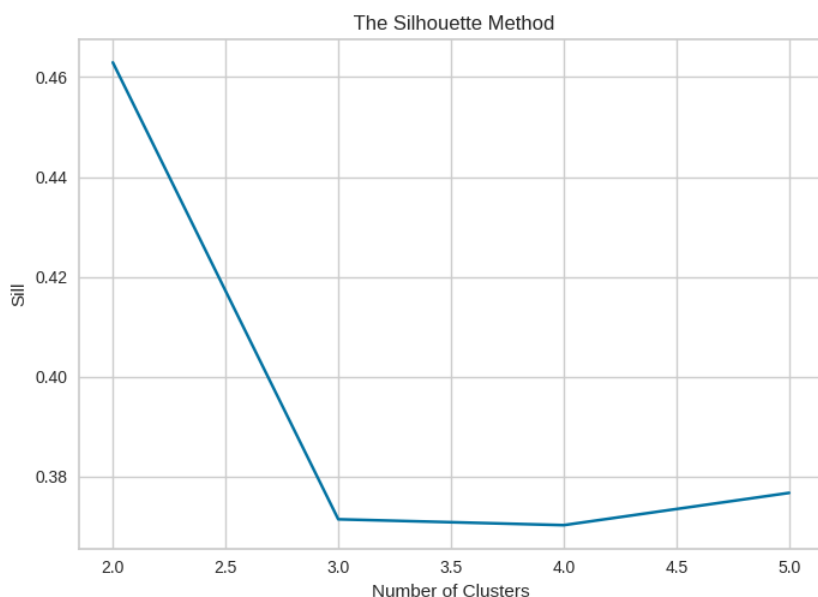
Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

YES

Clusters predicted by the elbow method are $k = 2$





Cluster 0

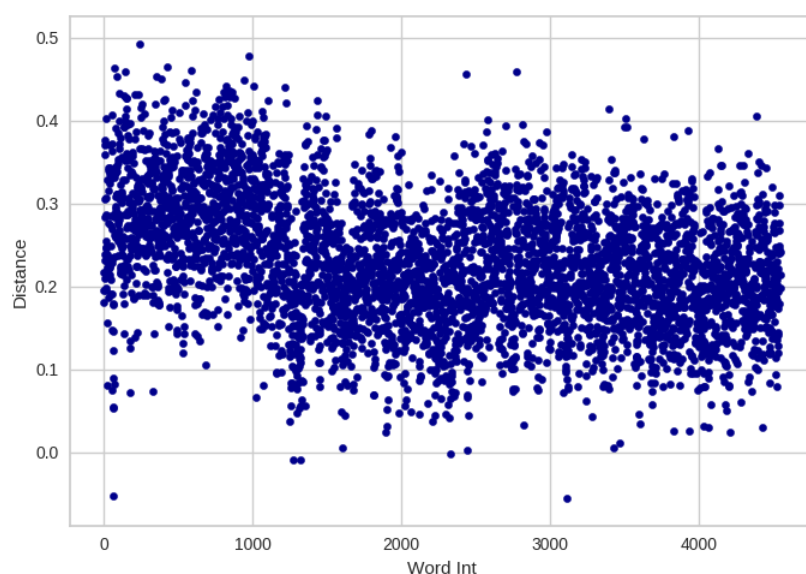
Implied Meaning: - “Pile of something. Stashing things one over the other”

Example Sentences:-

- 1) When the silo is filled or the **stack built**, a layer of straw or some other dry porous substance may be spread over the surface. In the silo the pressure of the material, when chaffed, excludes air from all but the top layer; in the case of the stack extra pressure is applied by weights in order to prevent excessive heating.
- 2) The chopped grass can then be dumped in at the top, to be drawn from the bottom in winter. This requires considerable effort to compress the **stack** in the silo to cure it properly. Again, the pit is covered with plastic sheet and weighed down with tires.
- 3) Before anaerobic fermentation starts, there is an aerobic phase in which the trapped oxygen is consumed. How closely the fodder is packed determines the nature of the resulting silage by regulating the chemical reactions that occur in **the stack**.
- 4) Virtanen was awarded 1945 Nobel prize in chemistry "for his research and inventions in agricultural and nutrition chemistry, especially for his fodder preservation method", practically inventing modern silage. Early silos were made of

stone or concrete either above or below ground, but it is recognized that air may be sufficiently excluded in a tightly **pressed stack**, though in this case a few inches of the fodder round the sides is generally useless owing to mildew.

Context Words: - tower, pile tip, chamber, heights, racks, closet, settling, stacking, head, hub, layers, funnel, columns, blocks, piles etc.



Threshold Plot for Cluster 0

Cluster 1

Implied Meaning: - “Data Structure”

Example Sentences: -

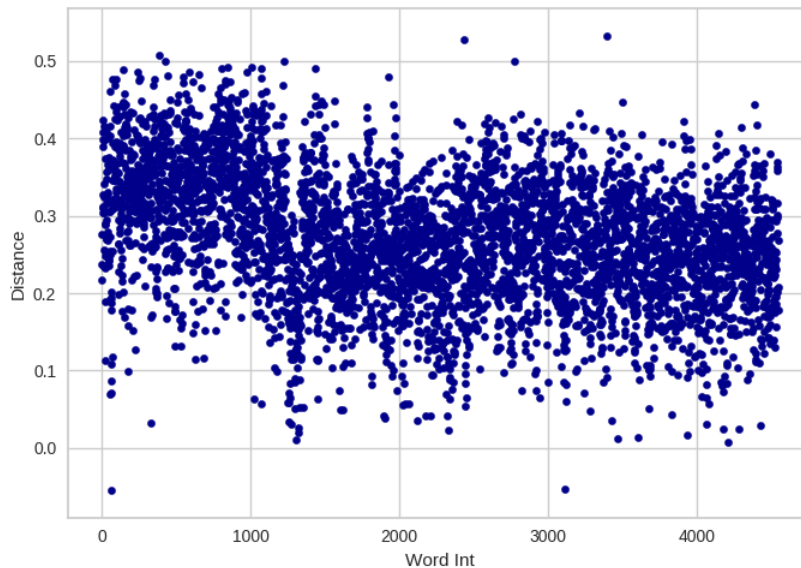
1) **Stack** is an abstract data type with a bounded(predefined) capacity. It is a simple data structure that allows adding and removing elements in a particular order. Every time an element is added, it goes on the top of the stack and the only element that can be removed is the element that is at the top of the stack, just like a pile of objects.

2) **Stack** is a LIFO (Last in First out) structure or we can say FILO (First in Last out). $push()$ function is used to insert new elements into the Stack and $pop()$ function is used to remove an element from the stack. Both insertion and removal are allowed at only one end of Stack called Top

3) For example, the abstract **data type stack** has the operations $new()$, $push(v, S)$ and $popOff(S)$, among others. These may be defined with the following axioms..

4) Since $push$ leaves **stack** non-empty, two operations undefined hence $invalid = \Lambda$. On hand, axioms and lack side effects imply $pushs, x = pusht, y x = = t$. As branches mathematics, customary assume **also stack** states whose existence proved axioms finite number steps. In abstract stack example above, rule means every stack finite sequence values, becomes empty stack Λ finite number pops.

Context Words: - array, stacked, loop, page, lists, heap, queue, switch, frame, list, buffer, cluster, data, structure, pop, push, pull, variable, function, compiler, processors, table, disk, query, call, definition, swap, sort etc.



Threshold Cluster Plot for Cluster 1



Section 3: - Examples of Overlapping-Clusters

Target word: - opaque

No of clusters: - 2

Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

2 out of 3 Clusters Make Sense. 3rd Cluster becomes redundant as it implies the same sense as implied in Cluster 2.

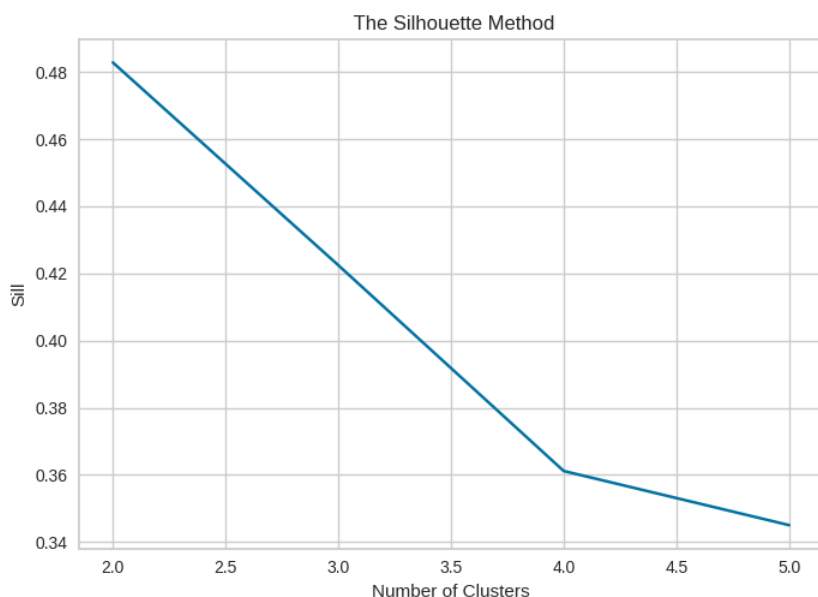
The clusters predicted by the elbow method are 3.

The clusters predicted by silhouette method are 2.

Sil Score for $K=2$ is 0.612

WCSS for elbow method for $K=2$ is 43 and the distortion score is 0.0253

Elbow Plot



Silhouette Plot

Cluster 0

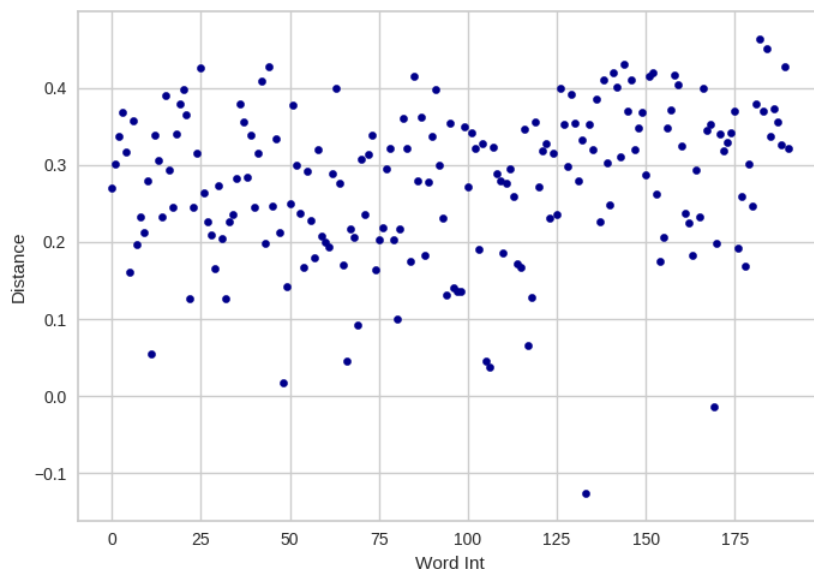
Implied Meaning: - “not transmitting or reflecting light or radiant energy; impenetrable to sight”

Example Sentences: -

- 1) For example, if the transmission of the barrier goes from **opaque** to transparent sharply as it moves from barrier to slit then this produces a wide diffraction pattern and consequently more crosstalk. If the transition is smoother then the diffraction will not spread so widely and less crosstalk will be produced.
- 2) For example, if the transmission of the barrier goes from **opaque** to transparent sharply as it moves from barrier to slit then this produces a wide diffraction pattern and consequently more crosstalk. If the transition is smoother then the diffraction will not spread so widely and less crosstalk will be produced.
- 3) If any **opaque** object is found between the surface and the light, the surface is in shadow and the light does not illuminate it. This recursive ray tracing added more realism to ray traced images.

4) In other traditional restorations, the porcelain is layered onto a metal substructure and often display colour brightness, an **opaque** "headlight", and dark oxide lines a "black line" in the vicinity of the gum line.

Context Words: - translucent, transparent, black, clear, dark, brightness, light, matt, masked, greyish, luminescent, spots, glass, mercury, stony, covered, brightly, scattering, sphere, screen, ivory, solid, reflection, reflective, clear, obscure etc.



Threshold Plot for Cluster 0

Cluster 1

Implied Meaning :- “A data type whose concrete data structure is not defined in an interface”

Example Sentences: -

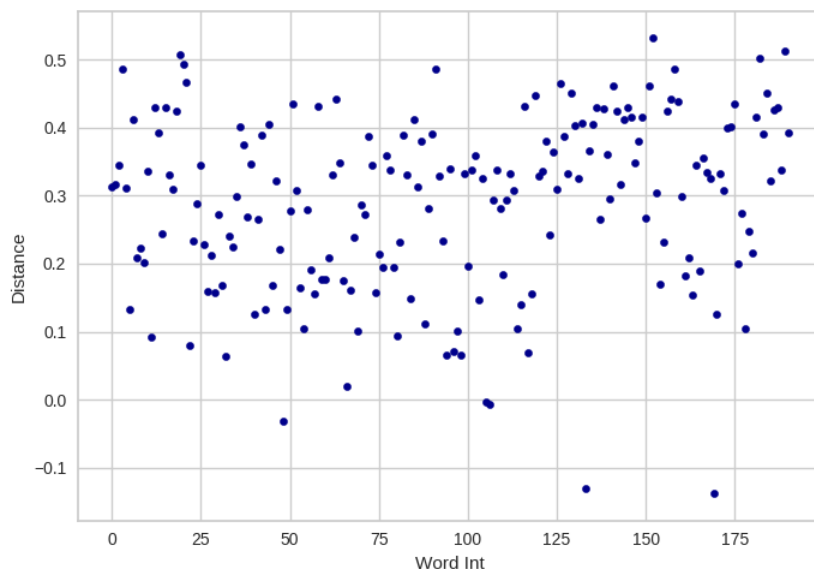
1) A resource handle can be an **opaque identifier**, in which case it is often an integer number (often an array index in an array or "table" that is used to manage that type of resource), or it can be a pointer that allows access to further information. Common resource handles include file descriptors, network sockets, database connections, process identifiers (PIDs), and job IDs.

2) Common resource handles include file descriptors, network sockets, database connections, process identifiers (PIDs), and job IDs. PIDs and job IDs are explicitly visible integers; while file descriptors and sockets (which are often implemented as a form of file descriptor) are represented as integers, they are typically considered **opaque**.

3) Doubly indirect handles have fallen out of favour in recent times, as increases in available memory and improved virtual memory algorithms have made the use of the simpler pointer more attractive. However, many operating systems still apply the term to pointers to opaque, "private" data structures—**opaque** pointers—or to indexes into internal arrays passed from one process to its client.

4) Arbitrary identifiers might lack metadata. For example, if a food package just says 100054678214, its ID may not tell anything except identity—no date, manufacturer name, production sequence rank, or inspector number. In some cases, arbitrary identifiers such as sequential serial numbers leak information (i.e. the German tank problem). **Opaque** identifiers—identifiers designed to avoid leaking even that small amount of information—include "really opaque pointers" and Version 4 UUIDs.

Context Words:- explicitly, visible, operators, blocks, hide, references, abstract, indirect, explicitly, encoded, manipulated, hidden, private, raw, concrete, inappropriate, alias, specified, abstraction etc.



Threshold Plot for Cluster 1



Target word: - Backup

No of clusters: - 2

Deterministic Method :- Elbow Method

Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

Only 2 out of the 3 clusters make sense. The 3rd Cluster becomes redundant as the implied meaning is same as that of in cluster 0.

Predicted clusters by elbow method = $k=3$

Predicted Clusters by silhouette method = $k=5$ (diff between sill score of $k=5$ and $k=2$ is 0.0275 which is very less and hence, the credibility of 5 clusters come int question.)

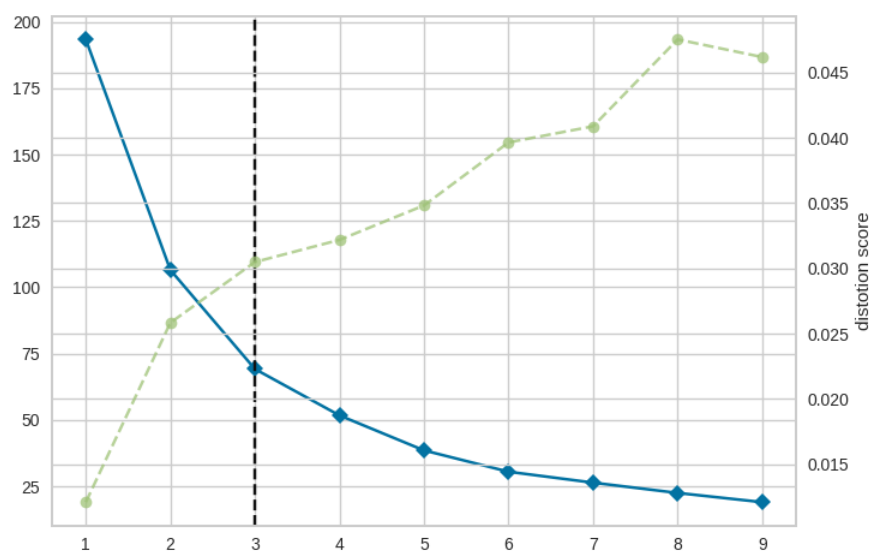
Cap method restricts the clusters to a value of 2.

For Elbow,

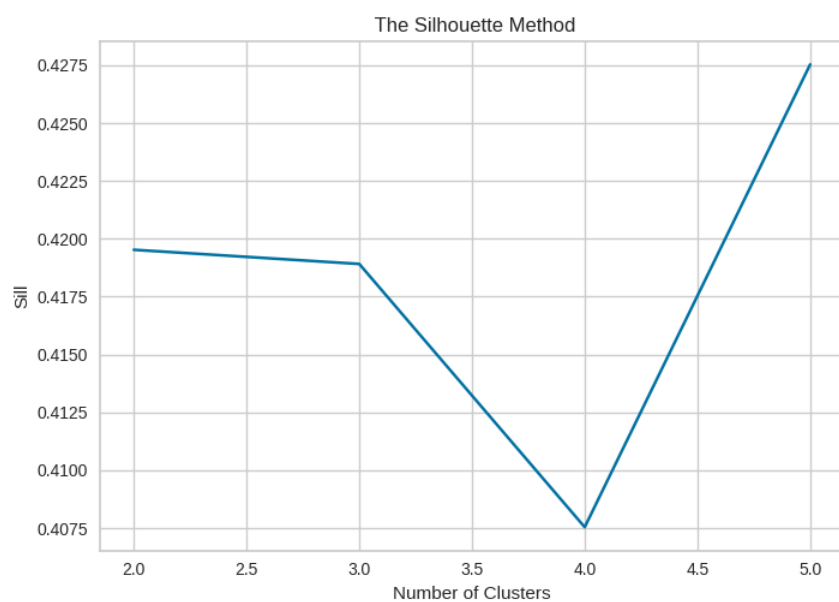
WCSS for $k=2$ is 110 and distortion score is 0.030

For Silhouette score,

Sil Score for $k= 2$ is 0.4200 and $k = 0.4275$



Elbow Plot



Silhouette Plot



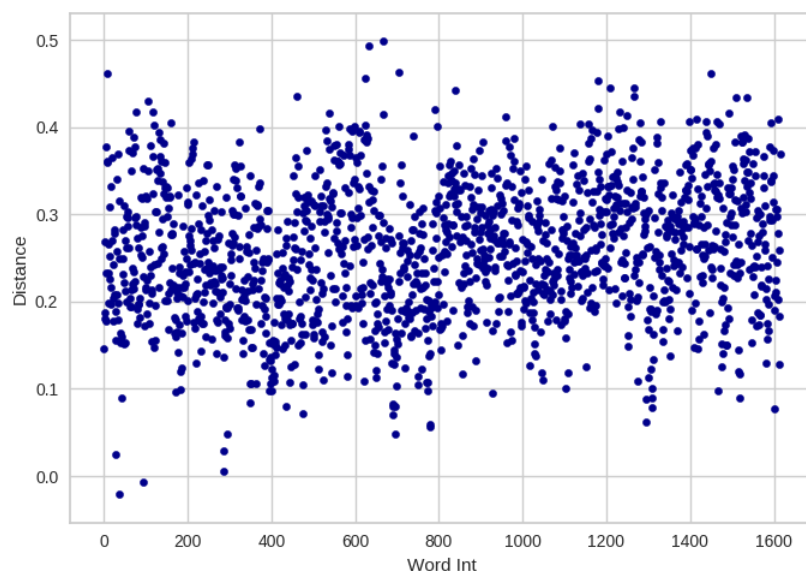
Cluster 0

Implied Meaning :- “Emergency support” (n.)

Example Sentences: -

- 1) Water-powered ejector pumps have a separate pump, float and check valve. The float controlling a **backup pump** is mounted in the sump pit above the normal high water mark. Under normal conditions, the main electric powered sump pump will handle all the pumping duties.
- 2) Many homeowners have inherited their sump pump configurations and do not realize that the pump discharges into the sewer. Usually hardwired into a home's electrical system, sump pumps may have a **battery backup**.
- 3) Since a sump basin may overflow if not constantly pumped, **a backup system** is important for cases when the main power is out for prolonged periods of time, as during a severe storm. There are generally two types of sump pumps—pedestal and submersible.
- 4) Aircraft also typically have dual "redundant" ignition systems which are nearly impossible to tune and time to produce identical ignition timing so using a lighter fuel that's less prone to autoignition is a wise "insurance policy". For the same reasons, those lighter fuels which are better solvents are much less likely to cause any "varnish" or other fouling on the **"backup" spark plugs**.

Context Words: - redundant, auxiliary, trigger, activate, emergency, support, temporary, timing, auto, reliable, exhaust, dual, storage, separate, secondary, primary, repeat, generators, failed, shortage, systems, operate, shift, mechanism, supplying, conditioning, supplying etc.



Threshold Plot for Cluster 0

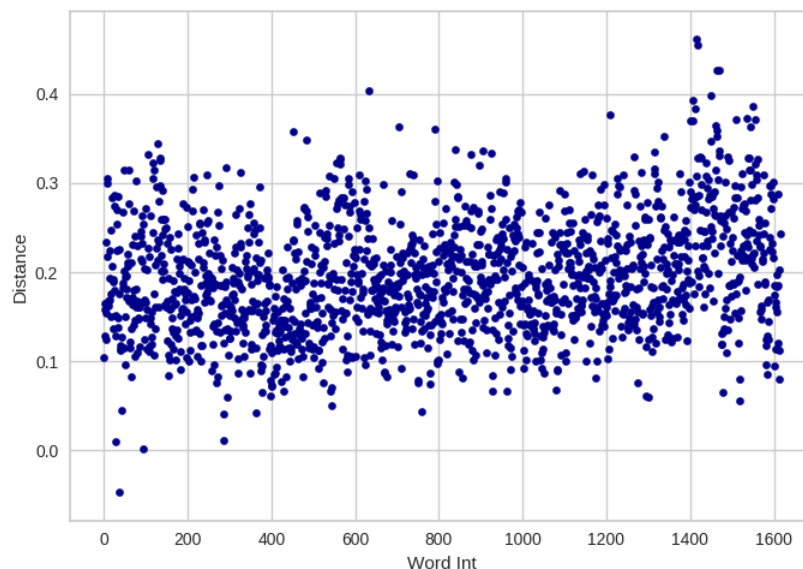
Cluster 1

Implied Meaning: - “(computer science) a copy of a file or directory on a separate storage device”

Example Sentences: -

- 1) To achieve this, a **backup** operation is done occasionally or continuously, where each desired database state (i.e., the values of its data and their embedding in database's data structures) is kept within dedicated backup files (many techniques exist to do this effectively).
- 2) Infrastructure as a service (IaaS) are online services that provide high-level APIs used to dereference various low-level details of underlying network infrastructure like physical computing resources, location, data partitioning, scaling, security, **backup** etc.
- 3) In many cases, the cryptosystem's structure involves back and forth communication among two or more parties in space (e.g., between the sender of a secure message and its receiver) or across time (e.g., cryptographically protected **backup** data). Such cryptosystems are sometimes called cryptographic protocols.
- 4) In information technology, a **backup, or data backup** is a copy of computer data taken and stored elsewhere so that it may be used to restore the original after a data loss event. The verb form, referring to the process of doing so, is "back up", whereas the noun and adjective form is "backup".[1] Backups can be used to recover data after its loss from data deletion or corruption, or to recover data from an earlier time.

Context Words: - copy, recovery, storage, protected, reproduce, restore, restore, update, secure, restart, dump, swap, maintenance, disaster, primary, copied, redundant, corrupted, accounts, drive, networking, archive, deployment, laptop, repairing, inspections, replicate, corruption, server, cloud, centers, hosting, protection etc.



Threshold Plot for Cluster 1

Target word: - Fork

No of clusters: - 3

Deterministic Method :- C-Elbow Method

Clustering Algorithm: - K-Means Clustering

Do the number of clusters make sense?

NO

Optimal No of Clusters :- 2

Clusters predicted by elbow method :- 4

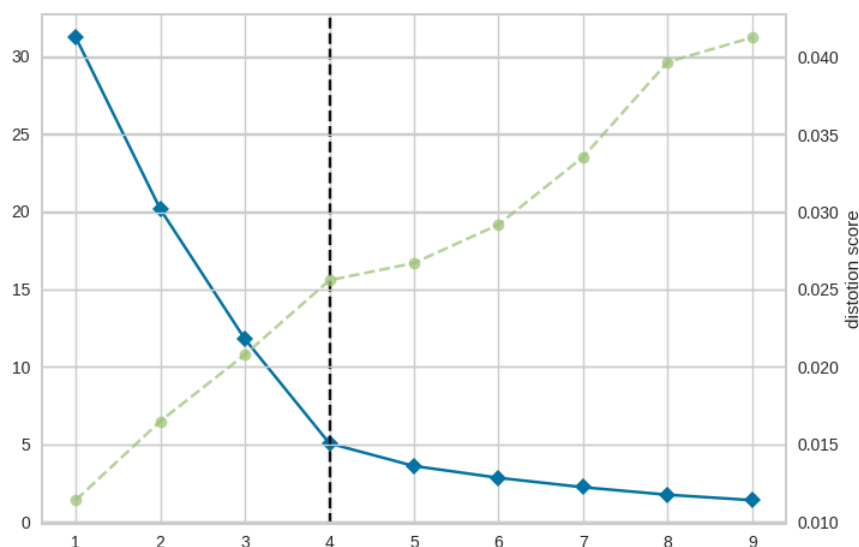
Clusters predicted by silhouette method: - 4

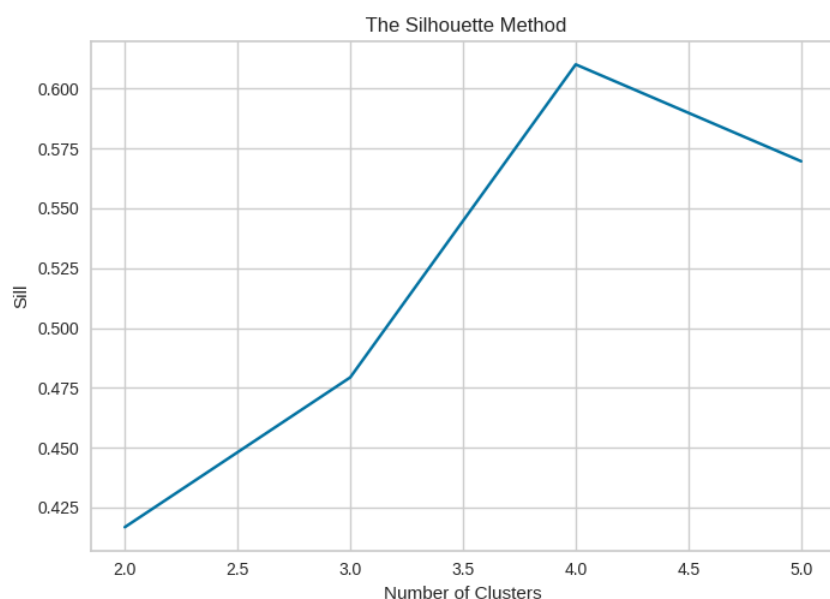
The no of clusters predicted seems to be ambiguous due to the lower frequency of occurrence of the word fork in the corpus cluster.

In the four clusters obtained by the cap method, the 2nd and the 3rd clusters hold the same contextual sense and the quantity of sentences collected in the third cluster are also very less as compared to the other two. It seems to be a false positive cluster predicted by the clustering algorithm.

A similar overlap can be observed between the sentences aggregated in Cluster 0 and Cluster 3. The sentences in cluster 3 are less than the ones in cluster 4 and the implication of the word *Fork* remains the same in both these clusters.

Though one key thing to note is that the two predicted clusters are crisp in their respective contexts and can be easily discerned from each other.





Cluster 0 and Cluster 3 (Both the clusters have the same implication of the word)

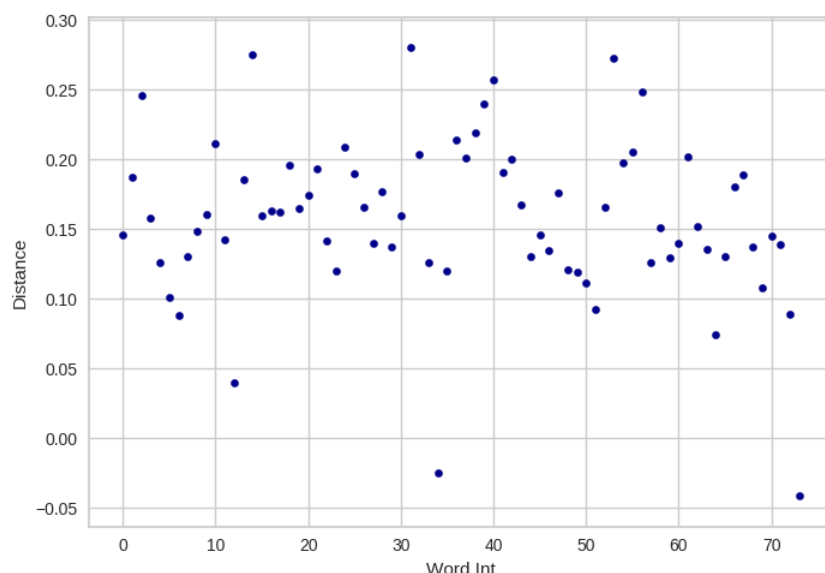
Implied Meaning: - “Divide into two or more branches so as to create a fork”

Example Sentences: -

- 1) Following pseudocode describes merge sort parallel recursion using **fork join keywords**:
`// Sort elements lo hi exclusive array A.algorithm mergesortA, lo, hi is lo+1 < hi // Two elements. mid := ⌊lo + hi / 2⌋ fork mergesortA, lo, mid mergesortA, mid, hi join mergeA, lo, mid, hi`
 This algorithm trivial modification sequential version parallelize well.
- 2) Its just-in-time compiler brings significant speed improvement CPython several libraries written C cannot used it. [112][113]Stackless Python significant **fork** CPython implements microthreads; use C memory stack, thus allowing massively concurrent programs. PyPy also stackless version.
- 3) On hand, empty trees simplify defining fixed branching factor: empty trees allowed, binary tree tree every node exactly two children, tree possibly empty. The complete sets operations tree must include **fork operation**.
- 4) Notable software packages, previously proprietary, open sourced include:Netscape Navigator, code became basis Mozilla Firefox web browsersStarOffice, became base OpenOffice.org office suite LibreOfficeGlobal File System, originally GPL'd, made proprietary 2001?, 2004 re-GPL'd.SAP DB, became MaxDB, distributed and owned MySQL ABInterBase database, open sourced Borland 2000 presently exists commercial product open-**source fork** FirebirdBefore

changing license software, distributors usually audit source code third party licensed code would remove obtain permission relicense.

Context Words: - join, merge, branch, forking, branching, git, release, sperate, tree, value, release, insertion, sort, stack, version, user, changes etc.



Cluster 1 and 2 (Both of the clusters have the same meaning)

Implied Meaning:- “Cutlery used for eating food”

Example Sentences: -

- 1) If soup is the first course, to the left of the dinner plate, moving clockwise, are placed a small salad fork to the left of the dinner plate; a large **dinner fork** to the left of the salad fork; a side plate above the forks.
- 2) The **dinner fork** is placed immediately left of the dinner plate; the salad fork is placed on the outer left side of the place setting.
- 3) When more courses are being served, place settings may become more elaborate and cutlery more specialised. Examples include fruit spoon or fruit knife, cheese knife, and **pastry fork**.

4) It is therefore a fairly common grip since it allows its user to hold all kinds of everyday objects such as a **fork** or a pen.

Context Words: - knife, knives, spoon, plate, cheese, glass, pastry, arm, napkin, food, cutler, course, spoon, glasses, plate, dinner, food, dishes, course, fruit, dairy, etc.

