

Assignment Problem:

CNN News corpus dataset: Link to download dataset -

https://drive.google.com/uc?export=download&id=0BwmD_VLjROrTHk4NFg2SndKcjQ

Pick up a sentences from the CNN News corpus dataset which should be related to the given 6 different sentences below:

1.Realistic — Realistic occupations frequently involve work activities that include practical, hands-on problems and solutions. They often deal with plants, animals, and real-world materials like wood, tools, and machinery. Many of the occupations require working outside, and do not involve a lot of paperwork or working closely with others.

2.Investigative — Investigative occupations frequently involve working with ideas, and require an extensive amount of thinking. These occupations can involve searching for facts and figuring out problems mentally.

3.Artistic — Artistic occupations frequently involve working with forms, designs and patterns. They often require self-expression and the work can be done without following a clear set of rules.

4.Social — Social occupations frequently involve working with, communicating with, and teaching people. These occupations often involve helping or providing service to others.

5.Enterprising — Enterprising occupations frequently involve starting up and carrying out projects. These occupations can involve leading people and making many decisions. Sometimes they require risk taking and often deal with business.

6.Conventional — Conventional occupations frequently involve following set procedures and routines. These occupations can include working with data and details more than with ideas. Usually there is a clear line of authority to follow.

Sample output after data collection will look like this

Realistic	Investigative	Artistic	Social	Enterprising	Conventional
Achieving an ideal is rarely, if ever, practical	Middle school students focus on core subject skills development	Emotion and social skills play two roles for an intelligent agent.	One of the main functions of most software metering programs is to keep track of the software usage statistics in an organization.	Analysis may often be carried out by engineers or programmers, rather than personnel with the domain knowledge to understand a client's needs properly.	There is considerable responsibility and autonomy and control or guidance of others is often required.
This is a good task for robots, not humans	The curriculum is designed so that skills and content taught in one subject are reinforced and developed in others	The simplest intelligent agents are programs that solve specific problems.	These tools are designed to bridge the communication gap between business users and the IT organization and also to allow applications to be 'test marketed' before any code is produced	This assists the IT departments in keeping track of licensed software, which is often from multiple software vendors.	Motivation has been shown to have roots in physiological, behavioral, cognitive, and social areas.

Once you create the dataset, your job is to combine the sentences of three each into a single sentences, along with the 3 labels (Permutation)

Sample output after permutation will look like this

Data	Label
Philosophical subject matter can be very abstract, complex and difficult for unfamiliar readers to understand. Social skills training teaches clients skills to access reinforcers and lessen life punishment. These three classes are somewhat hierarchical in nature, in terms of the level of conscious thought related to their use.	Realistic Investigative Social
In this the word Shilpa (sculptural) Ratna (Gems) is used as a broad term embodying artistic forms that either uses the body as a medium of expression (like Dance, Drama or Dance-Drama) or that which represents the body as an expression (like Sculptures and mural arts). Being able to identify these sub-structures within a network can provide insight into how network function and topology affect each other. Robot resembling the shape or form of a human.	Artistic Enterprising Conventional
The style can vary considerably between schools; some are very similar, yet others are very distinct. Hardware- and software-specific information is often stored in configuration files in specified locations (e.g. The roles are but three; a sign of an object leads to one or more interpretants, and, as signs, they lead to further interpretants.	Realistic Social Conventional

Then next step is to train the dataset which has combination of sentences with 3 Labels to produce a score

Sample output will look like given below

Data	Artistic	Conventional	Enterprising	Investigative	Realistic	Social
I want to become a doctor and service people	0.001909528	0.001688219	0.0637789	0.9866151	0.97878575	0.9988432
I love teaching and want to become a teacher like my father	0.44125783	3.03E-05	0.573817	0.9850395	0.6913145	0.9878357
I want to be a successful Business man. I want to take my family business to the next level. I think Business man are the most financially stable and live a happy life.	0.007018087	0.023182722	0.99986064	0.020278124	0.9828431	0.7469797
I want to become a nurse because our country needs more people who are ill and fight with Covid Situation	0.003550061	0.002616516	0.9653855	0.16371611	0.8496932	0.99972266
I want to become a banker because I can make more money and achieve my dreams.	0.004014205	0.006953222	0.99927396	0.046978783	0.99699605	0.79196095

Kindly share us the assessment submission to the email id suhas@sahicareer.com .

Let us drop a mail if you want more information