

## RULES FOR POS TAGGING NLP

### VERB

- If the sentence contain 'Na' than word after that is said to be 'Verb' and vice-versa
- In the present tense, Gujarati verbs typically end with "-યો" ("-yo") for singular subjects and "-યું" ("-yun") for plural subjects.
- The future tense is formed by adding auxiliary verbs such as "કરીશ" ("karish," will do) or "હશે" ("hashe," will be) before the main verb. The infinitive form of a verb in Gujarati is often used to express purpose and is formed by adding "-વું" ("-vum") to the verb stem
- Gujarati has a passive voice, which is formed by adding the passive marker "-વાયો" ("-vāyo") to the verb stem
- To form questions, auxiliary verbs like "શું" ("shum," do) are often used at the beginning of a sentence.

### ADVERB

- many adverbs are formed by adding the suffix "-માં" ("-mām") to adjectives.
- adverbs are generally placed immediately before the verb they modify.
- you can add the suffix "-થી" ("-thī") to the adverb.
- you can add the suffix "-થી" ("-thī") to the adverb.
- you can add the suffix "-તા" ("-tama") to the adverb.
- Adverbs typically comes before the verb

### ADJECTIVE

- Adjective mostly comes after or before noun in the statement.

## DETERMINANT

- The determinant is always statement independent , that is some word are Already determinant from its definition, irrespective of statement so no rules required for its pos tagging. આ (ā) આ પ્રકારનો (ā prakārno) મારી (mārī)

## PREPOSITION

- The preposition is always statement independent , that is some word are Already preposition from its definition, irrespective of statement so no rules required for its pos tagging. Ex: પર (par) ને (ne)

## NOUN AND PRONOUN

- Here in grammar for any language, one thing is common that noun and pronoun are universal so the pos tagging of noun and pronoun is totally dependent on the dataset or tag set created, there is no logic to describe a particular noun or pronoun for coding purpose .

## NUMBER

- Here, the number (numerical visual ) or written numbers (like લાખ, હજાર,કરોડ, સો will be pos tagged based on their appearance in the input statement whether it is numerical like 1000,1500,11:23.1254 etc or written numbers.

## PUNCTUATION

- we will give a string in if statement in rule based approach which will include full stop, comma, quotation, question mark, exclamatory statement, hyphen , semi colon , colon , slashes. And removing the special characters like @\$%&(){}[]\*+.-.

**20SE02ML009 JAY CHAUHAN**

**20SE02ML034 MANAN PATEL**

**20SE02ML036 VEDANT RAJPUROHIT**