Difficulty Level Prediction of a Question Paper Using Naive Bayes Classifier

Thipireddy Rishith Reddy[1], D.Sanjana Reddy[2], Manoj Boddu[3], Boda Bhagya Rishiroop[4],R.Rajkumar[5]
[1,3,4,5]School of Computer Science and Engineering, Vellore Institute of Technology
[2] Sreenidhi Institute of Science and Technology,Hyderabad

Abstract: The Naïve Bayes classifier is one of the most straightforward ways to deal with the grouping undertaking that is as yet equipped for giving sensible exactness. Bayesian induction, of which the gullible Bayes classifier is an especially straightforward illustration, depends on the Bayes decide that relates restrictive and negligible probabilities. Implementing such Naïve Bayes Algorithm to classify based on a decision attribute which is formed by mining the response dataset from several participants and assigning a class such as 'Tough', 'Medium', 'Easy'. And following this procedure is the approach that we have used in this project. So, the question paper can predict the difficulty level of the question paper for which he is going to set using this procedure. We have successfully encountered our problem statement of Difficulty Level Prediction of a Question Paper through Classification Algorithm and managed to master the topic of Opinion mining and Naïve Bayes Classifiers along.

Keywords: Opinion Mining, Naïve Bayes, Decision attribute, Categorical variable, Prediction.

1. Introduction

Frequently alluded to as Emotion AI, Opinion mining has been turning into an unmistakable aspect of our new digitalized life. To characterize it in proper words, it can expressed as the utilization of characteristic language preparing, text examination, computational etymology, and biometrics to efficiently distinguish, separate, measure, and study full of feeling states and emotional data. To talk when all is said in done, conclusion mining intends to decide the disposition of a speaker, author, or other subject regarding some point or the general logical extremity or passionate response to a record, intelligent, or occasion. The aura may be a judgment or appraisal (see assessment theory), loaded with feeling state (as it were, the energetic state of the maker or speaker), or the normal eager correspondence (as it were, the excited effect arranged by the maker or examiner).

This procedure of extricating human feelings from advanced logs is by and large broadly applied to the voice of the client materials, for example, audits and study reactions, on the web and web-based media, and medical care materials for applications that extend from promoting to client assistance to clinical medication. In our venture we have ventured forward to utilizing this strategy to separate the feeling of an understudy concerning an inquiry paper he/she endeavored by investigating the remark segment of a review reaction that the understudy filled to express his origination of the trouble level of that paper as 'Simple' or 'Medium' or 'Extreme'. Further, we have abused the idea of Naïve Bayes Classification through a calculation which utilizes the above removed feeling of various understudies to be filled as tuples of a choice characteristic used to order and foresee the degree of trouble of the inquiry paper in respect. Directly, Bayesian classifiers are verifiable classifiers. They can predict class enlistment probabilities, for instance, the probability that a given model, has a spot with a particular class. Bayesian classifier relies upon Bayes' theory. Guiltless Bayesian classifiers expect that the effect of a quality regard on a given class is liberated from the assessments of exchange properties. This doubt is called class unforeseen independence. It is made to revise the figuring included and, in this sense, is seen as "credulous". Innocent Bayes classifiers are extraordinarily flexible, requiring different boundaries direct in the amount of variables (features/markers) in a learning issue. Most prominent likelihood getting ready should be conceivable by evaluating a closed casing verbalization, which takes straight time, rather than by exorbitant iterative gauge as used for some various kinds of classifiers.

Naïve Bayes Theorem

Let X = {x1, x2, . . . , xn} be a sample, whose parts speak to values made of an arrangement of n qualities. In Bayesian terms, X is considered "proof". Give H a chance to be some theory, for example, that the information X has a place with a particular class C. For grouping issues, we will likely decide P (H|X), the likelihood that the theory H holds given the "proof", (i.e. the watched information test X). As such, we are searching for the likelihood that example X has a place with class C, given that we know the quality portrayal of X.

P (H|X) is the a posteriori likelihood of H conditioned on X. For case, assume our information tests have characteristics: age and salary, and that example X is a 35-year-old client with a pay of $40,000. Assume that H is the theory that our client will purchase a PC. At that point P(H|X) is the likelihood that client X will purchase a PC given that we know the client's age and salary. Converse, P (H) is the from the earlier likelihood of H. For our illustration, this is the likelihood that any given client will purchase a PC, paying little respect to age, salary, or any other data. The a posteriori likelihood P(H|X) depends on more data (about the client) than the from the earlier likelihood, P(H), which is free of X. Essentially, P(X|H) is the a posteriori likelihood of X adapted to H. That is, the likelihood a client X, is 35 years of age and wins $40,000, given that we know the client will purchase a PC. P(X) is the

ISSN No : 1006-7930

from the earlier likelihood of X. In our case, the likelihood a man from our arrangement of clients is 35 years old and procures $40,000.

According to Bayes' theorem, the probability that we want to compute P (H|X) can be expressed in terms of the probabilities P(H), P(X|H), and P(X) as

$$\mathbf{P(H|X) = (P(X|H).P(H))/P(X)} \text{ ------ eq(1)}$$

and these probabilities may be estimated from the given data.

2. Literature review

Sherica Lavinia Menezes and Geeta Varkey proposed a framework which predicts the missing things in view of the past data and recommends the same to the clients. To satisfy this undertaking the framework utilized grouping systems preceding the forecast procedure. The upside of utilizing characterization/grouping is that the expectation is done at a more elevated amount of deliberation and the cost of managing age in affiliation administer mining is limited. Out of the different choices accessible, Naive Bayes classifier is decided in order since this classifier will function admirably for extensive informer indexes and in a moderately easy to actualize. The various leveled grouping instrument is decided for bunching. The information extraction is finished utilizing a robotized information extraction instrument web Harvest. The information structure decided for charts is a Hash List which is a mix of the hash table and connected rundown and is appeared to be an effective information structure for speaking to a diagram. **J. Read** utilized Twitter gushing information gave by Firehouse, which gave all messages from each client continuously. They examined with quick incremental strategies that were good to manage information streams: stochastic inclination plunge, multinomial innocent Bayes and the Hoeffding tree. In this way they inferred that SGD-based model, utilized with a reasonable learning rate was the best. **Ruchi Mehra, Mandeep Kaur Bedi** presented an analysis of opinion conduct of Twitter information. The proposed work uses the guileless Bayes and fluffy Classifier to order Tweets into constructive, contrary or neural conduct of a specific individual. They introduced trial assessment of the dataset and characterization comes about which demonstrated that joined proposed technique is more effective as far as Accuracy, Precision and Recall.

Theresa Wilson, Janyee Wiebe et al., proposed another way to deal with state level assessment, investigation can be that initially decides if an articulation is impartial or polar and afterward disambiguates the extremity of the polar articulations. With this approach, the framework can consequently recognize the relevant extremity for an extensive subset of feeling articulations, accomplishing comes about that are fundamentally superior to pattern.

Simranjeet kour Bindra, Akshay Girdhar et al., discussed that the question paper generation is a manual approach, prompting ineffective now and again attributable to predisposition, redundancy and security concerns. The present paper shows a programmed method of question paper gather which can be adjusted, streamlined, synchronized and secured. Each assignment done by this framework is programmed, with the end goal that putting away space, inclination and security isn't a fear any more. Prior, the inquiry paper was created by concerned subject instructor physically and was extremely tedious, labor was required and once in a while the inquiry paper needed precision. Result Based Education (OBE) assigns what understudies will know and be savvy to do, as they progress in a program. **Dominic Seyler** et al., proposed an Automated question generation for quality control in human calculation errands. The issue of producing question things from ontologies has as of late increased much consideration in the software engineering group. This is principally because of the utility of the created inquiries in different instructive and expert exercises, for example, student evaluations in eLearning frameworks, quality control in human computational errands and, extortion location in crowdsourcing stages to give some examples. **Tahani Alsubait** et al., proposed an Ontology-based multiple-choice question generation. A Traditionally, question age (QG) approaches have, to a great extent concentrated on recovering inquiries from crude content, databases and other non-semantic based information sources. Nonetheless, since these sources don't catch the semantics of the space of talk, the produced questions can't be machine-handled, making them less employable in a large number of this present reality applications. For instance, the addresses that are produced from crude content are reasonable just for dialect learning errands. Utilizing semantics-based information sources in QG have different points of interest, for example, in ontologies, we demonstrate the semantic connections between area substances, which help in producing significant and machine-processable inquiries ontologies empower standard thinking and questioning administrations over the learning, giving a system for creating questions all the more effectively. An early push to distinguish factors that could possibly anticipate the trouble level was by Seyler et al., They have acquainted a technique with the group an inquiry as simple or hard by finding the highlights of the comparable inquiry elements in the Linked Open Data (LOD). Highlight esteems for the characterization assignment are gotten in light of the availability of the inquiry substances in the LOD.

Xinming An et al., proposed an Item response theory. This hypothesis was first proposed in the field of psychometrics, later, the hypothesis were utilized broadly in instructive research to adjust and assess questions things in the overall examinations, for example, the Scholastic Aptitude Test (SAT) and

Graduate Record Examination (GRE). **E.V Vinu** et al., proposed an Automated generation of assessment tests from domain ontologies. The creators have considered all the conceivable non specific inquiry designs that are helpful in producing basic real issues. They have additionally proposed strategies for choosing space significant resultant tuple55s/ inquiries from leading area related evaluations. **Gulijers JTM** et al., proposed a five-dimensional system for real appraisal. The best possible outline of appraisal organizations will clearly drive methods for an understudy's way to deal with learning. Improper plan of appraisal organizations may prompt undesirable results of the abilities and sorts of patient care. **Epstein RM** et al., created Assessment in restorative instruction. For as long as a couple of decades, numerous restorative instruction programs and permitting experts either at undergrad level or postgraduate level have dispensed enormous endeavors to guarantee the legitimacy of appraisals and competency of students.**Miller G** et al., proposed the assessment of clinical skills/competence/performance. E evaluation arrange has its points of interest and impediments relying upon the appraisal plan. The best appraisal technique must meet five criteria which incorporate dependability, legitimacy, agreeableness, plausibility and instructive effects on learning and practice. Mill operator arranged appraisal techniques into four classifications which incorporate knows (i.e. surveying learning), knows how (i.e. surveying capacity to apply learning inside its specific situation), demonstrate how (i.e. surveying learners' execution in recreated condition) and does (i.e. surveying students' execution in real condition).in actual environment).

.

3. Problem Statement and Methodology

In this paper, a solution to ‘Difficulty level prediction of a question paper’ has been formulated using a Naïve Bayes Algorithm.

i) A questionnaire shall be prepared using google forms and circulated among students who have answered a particular question paper. The questionnaire will ask the students to fill the difficulty level of each question in the question paper and a few other details about their CGPA. The questionnaire will also contain a free text area to comment about the entire question paper as a whole. The sentiment analysis of this text area will be taken as the decision attribute of that particular student.

ii) On opinion mining the comment section of the responses, we have distributed each entry into one of three categories –‘Easy’, ‘Medium’, ‘Tough’ under a single attribute. This becomes the decision attribute of the tuple for our classifier.

iii) Now Naïve Bayes is performed on a new tuple that we will input to get the prediction whether that particular paper in talk is hard or not.

iv) The question paper on any subject will have 5 Questions where the dataset is collected according to the single question paper on a particular subject.

Given a dataset will have a total of 7 attributes, out of which all are categorical variables. The categorical features are CGPA, Qsno 1, Qsno 2, Qsno 3, Qsno 4, Qsno 5, Overall remark. The attribute CGPA will have three categories Average, Above average and Below average. The attributes from Qsno 1 to Qsno 5 will have three categories Hard, Medium, Easy. The Final attribute which is a decision attributed known as Remark will also have three classes 'Tough', 'Medium', 'Easy'.

$$\mathbf{P(X|Ci) = P(x1|Ci)*P(x2|Ci)*\ldots*P(xn|Ci)} \quad - \quad \mathbf{eq(2)}$$

v) We can easily estimate the probabilities P(x1|Ci), P(x2|Ci), .. , P(xn|Ci) from the training tuples. Here *xk* refers to the value of attribute *Ak* for tuple ***X***.

vi) In order to predict the class label of ***X***, P(**X**jCi)P(Ci) is evaluated for each class *Ci*. The classifier predicts that the class label of tuple ***X*** is the class *Ci* if and only if

$$\mathbf{P(X|Ci)P(Ci) > P(X|Cj)P(Cj) \text{ for } 1 <= j <= m, j != i} \quad - \quad \mathbf{eq(3)}$$

In other words, the predicted class label is the class *Ci* for which P(**X**|Ci)P(Ci) is the Maximum

Architecture Flow:

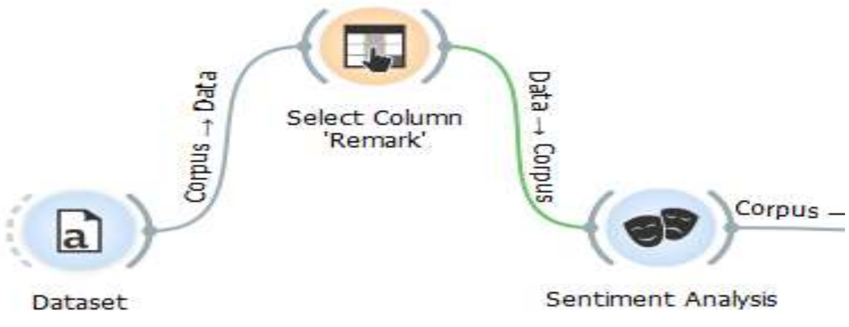

Fig 1. The Decision attribute column is subjected to Sentiment analysis

Figure 1 depicts that the collected dataset will have the unprocessed ‘Remark’. It is selected as selected column and Sentiment analysis is performed to categorize into three classes ‘Tough’, ‘Medium’, ‘Easy’.

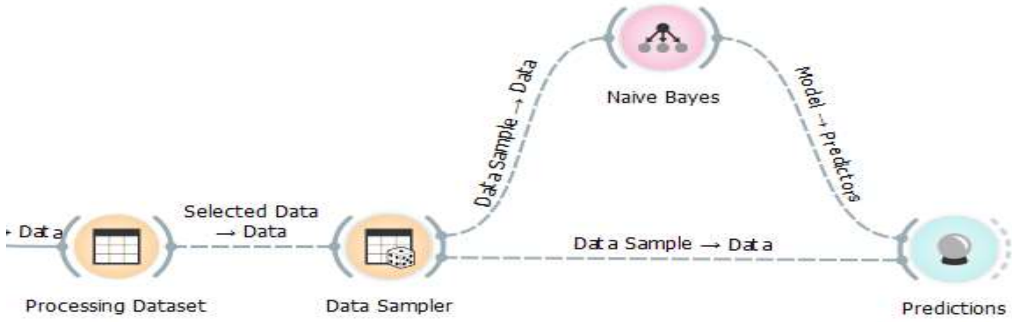

Fig 2. The Prediction phase

Figure 2 depicts that after the sentiment analysis phase, the processed data set used in predicting the predictions.

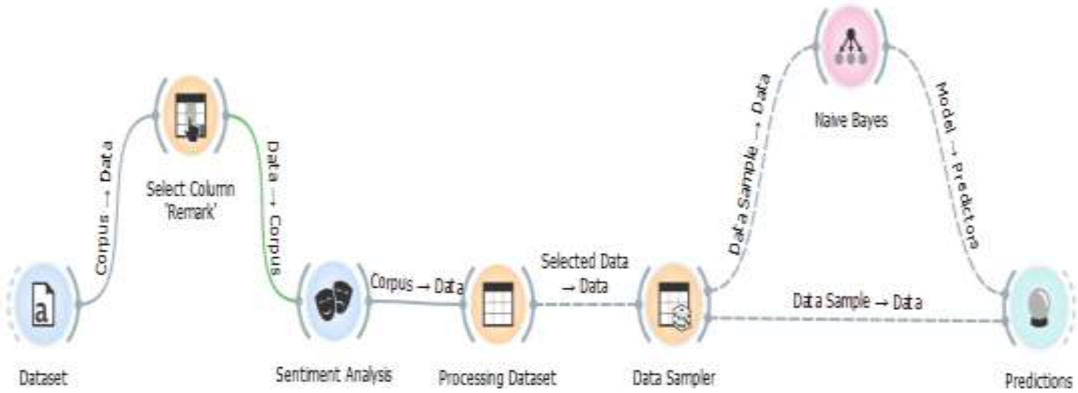

Fig 3. The Overall Architectural Flow

3. Results and Discussion

The question paper of any subject will have 5 Questions where the dataset is collected according to the single question paper of a particular subject and that sample questions is used to analise the remarks of that questions. The questions contains various transactions, customer’s Id, age, income, credit cards and class. The dataset we have collected from the class of 65 students, rating levels for each question is taken by the student as shown below.

ISSN No : 1006-7930

Username	CGPA:	Name:	Question.	Question.	Question.	Question.	Question.	Overall Remark on Paper (Please be as elaborate as possible)
nikhil.laki	9.15	Nikhil	medium1	easy2	hard3	hard4	medium5	Paper is tough
vynathey.	9.62	Sree Vyna	easy1	medium2	hard3	medium4	easy5	Paper is not too tough
gkailashn	9.22	kailash	medium1	medium2	medium3	hard4	hard5	Tough
b.praveer	9.32	Praveen R	easy1	hard2	hard3	easy4	medium5	Paper is very lengthy to attempt
kailasasar	8.6	sandeep	hard1	hard2	hard3	hard4	hard5	Paper is very very hard
yeshwant	8	Yeshwant	easy1	easy2	hard3	hard4	easy5	Paper is medium
kamisetty	8.4	Yeshwant	easy1	easy2	easy3	easy4	easy5	Only numericals are given, it was easy to solve.
nikhilesh.	8.34	Chamarth	medium1	easy2	easy3	medium4	medium5	It was a good paper, clear questions
jaya.krishi	8	Jayakrishr	easy1	medium2	hard3	hard4	hard5	Paper is hard
ppravesh.	9.13	Pravesh Ji	easy1	easy2	medium3	easy4	hard5	I found only one question hard, rest were easy
karichetib	8.29	Bhargav	medium1	medium2	medium3	hard4	hard5	The questions were fine, could attempt all of them
himani.jai	8.8	Himani Ja	medium1	medium2	easy3	hard4	easy5	The questions were medium, i did well
geetanjali	9.8	Geetanjal	medium1	easy2	easy3	easy4	easy5	Very challenging paper. Amazing.
nikhil.laki	9.15	Nikhil	easy1	easy2	hard3	easy4	medium5	Paper is easy but very lengthy.
vadlaman	8.82	Shashank	hard1	easy2	easy3	easy4	easy5	Pretty normal
venkata.s	8.75	Saikrishna	medium1	medium2	medium3	medium4	medium5	Good Paper is easy but very lengthy.
rashigaikv	8.9	Rashi Gaik	medium1	hard2	hard3	medium4	easy5	Difficulty level was high.
myself@g	6.4	Me	hard1	hard2	medium3	easy4	medium5	really tough to attempt
anukritipa	8	Anukriti	medium1	medium2	easy3	medium4	easy5	The
tarangven	10	Tarang Ve	easy1	easy2	easy3	easy4	easy5	easy paper
venkata.s	8.75	P. V. Saikr	easy1	easy2	medium3	hard4	hard5	Average paper
jzz.sprous	8.87	jasprit ka	medium1	medium2	medium3	medium4	medium5	Paper was predictable and medium
satyamch	6	Satyam	hard1	hard2	hard3	hard4	hard5	Very very hard
abhilasha	9.4	Abhilasha	medium1	medium2	easy3	medium4	easy5	Paper was fairly easier than expected. A little more thought provoking questions need to be put.
priyam25	8.12	Priyam Ka	hard1	hard2	hard3	hard4	hard5	Hard paper
paruljaisv	9.8	Parul Jais	easy1	easy2	easy3	easy4	easy5	It was easy..

Fig 4: Dataset

The Sentiment analysis is performed on the decision attribute ‘overall remark on paper’ and the continous attribute ‘CGPA’ is broadly classified into 3 categorical variables such as ‘above avg’, ‘below avg’, ‘avg’. ‘above avg’ are those with CGPA > 9.0, ‘avg’ are those with CGPA between 8.0 and 9.0 . ‘belowavg’ are those with CGPA < 8.0 .

CGPA:	Question.	Question.	Question.	Question.	Question.	Overall Remark on Paper (Please be as elaborate as possible)
aboveAvg	medium1	easy2	hard3	hard4	medium5	tough
aboveAvg	easy1	medium2	hard3	medium4	easy5	medium
aboveAvg	medium1	medium2	medium3	hard4	hard5	tough
aboveAvg	easy1	hard2	hard3	easy4	medium5	tough
avg	hard1	hard2	hard3	hard4	hard5	tough
avg	easy1	easy2	hard3	hard4	easy5	medium
avg	easy1	easy2	easy3	easy4	easy5	easy
avg	medium1	easy2	easy3	medium4	medium5	medium
avg	easy1	medium2	hard3	hard4	hard5	tough
aboveAvg	easy1	easy2	medium3	easy4	hard5	easy
avg	medium1	medium2	medium3	hard4	hard5	medium
avg	medium1	medium2	easy3	hard4	easy5	easy
aboveAvg	medium1	easy2	easy3	easy4	easy5	easy
aboveAvg	easy1	easy2	hard3	easy4	medium5	tough
avg	hard1	easy2	easy3	easy4	easy5	medium
avg	medium1	medium2	medium3	medium4	medium5	tough
avg	medium1	hard2	hard3	medium4	easy5	tough
belowavg	hard1	hard2	medium3	easy4	medium5	tough
avg	medium1	medium2	easy3	medium4	easy5	medium
aboveAvg	easy1	easy2	easy3	easy4	easy5	easy
avg	easy1	easy2	medium3	hard4	hard5	medium
avg	medium1	medium2	medium3	medium4	medium5	medium
belowavg	hard1	hard2	hard3	hard4	hard5	tough
aboveAvg	medium1	medium2	easy3	medium4	easy5	easy
avg	hard1	hard2	hard3	hard4	hard5	tough
aboveAvg	easy1	easy2	easy3	easy4	easy5	easy

Fig 4a: The dataset after performing necessary actions.

Now the question paper setter will give the tuple to predict the difficulty level.

```
Enter the tuple
aboveavg,medium1,medium2,hard3,medium4,easy5
class probability of medium is 0.41379310344827586
split data is ['aboveavg', 'medium1', 'medium2', 'hard3', 'medium4', 'easy5']
class probability of tough is 0.3793103448275862
split data is ['aboveavg', 'medium1', 'medium2', 'hard3', 'medium4', 'easy5']
class probability of easy is 0.20689655172413793
split data is ['aboveavg', 'medium1', 'medium2', 'hard3', 'medium4', 'easy5']
Medium
>>> ================================ RESTART ================================
Enter the tuple
belowavg,medium1,medium2,medium3,medium4,medium5
class probability of medium is 0.41379310344827586
split data is ['belowavg', 'medium1', 'medium2', 'medium3', 'medium4', 'medium5']
class probability of tough is 0.3793103448275862
split data is ['belowavg', 'medium1', 'medium2', 'medium3', 'medium4', 'medium5']
class probability of easy is 0.20689655172413793
split data is ['belowavg', 'medium1', 'medium2', 'medium3', 'medium4', 'medium5']
Medium
>>> ================================ RESTART ================================
```

Fig 5: Output

5. Conclusion

The inquiry paper setter can utilize this framework to know the assessment of an understudy before really setting the inquiries, so the paper setter can adjust as per the understudy's scholarly conduct. This will fill in as a stride ahead in the training framework. Here, We have effectively experienced our concern articulation of Difficulty Level Prediciton of a Question Paper through Classification Algorithm and figured out how to ace the subject of Opinion mining and Naïve Bayes Classifiers along.

References

1. Bindra, S. K., Girdhar, A., & Bamrah, I. S. (2017, October). Outcome based predictive analysis of automatic question paper using data mining. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)* (pp. 629-634). IEEE.
2. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford, 1*(12), 2009.
3. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations Trends Information Retrieval 2 (1–2): 1–135.
4. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.
5. Lowd, D., & Domingos, P. (2005, August). Naive Bayes models for probability estimation. In *Proceedings of the 22nd international conference on Machine learning* (pp. 529-536).
6. Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 347-354).