# Categorizing the Content of GitHub README Files

BHAVYAI GUPTA, University of Calgary, Canada

KAYODE AWE, University of Calgary, Canada

MICHAEL MAN YIN LEE, University of Calgary, Canada

## SUMMARY OF CONTRIBUTIONS:

### 1.1 Data Collection

**Bhavyai** developed script that downloads the README.md file from random GitHub repositories using GitHub API. The script was initially written as `.py` script, and then later modified to run on Databricks. Once few hundred README.md files are downloaded, **Michael** selected 86 files with a total of 1000 sections for manual annotation. Each of the author of this project annotated all the 1000 sections. **Kayode** is the facilitator to make sure we reach an agreement of what final manual annotations should be used as well as calculating the statistics such as the Cohen Kappa analysis to show our overall process. **Michael** is responsible for extracting the sections from README files and to feed the new annotated data back to the database for modelling and analysis.

### 1.2 Coding

**Bhavyai** is responsible for developing script to download all new README files needed for this project. **Michael** is responsible for the initial draft of modifying the original code from the author to work on databricks.

**Kayode** is responsible for creating the notebook for comparing the results of the old and old plus the new annotated data using the original research models.

**Michael** is responsible for creating the notebook for comparing the results of the old and new annotated datasets using previously unexplored models.

**Bhavyai** is responsible for creating the notebook for hyperparameter optimization of the models.

### 1.3 Writeup

The breakdown of the report writeup work can be found in the table below:

| Section | Resource |
| --- | --- |
| **Page 1 contents** | Michael |
| **Abstract/Introduction/Conclusion** | Bhavyai |

| How was the new data labeled/collected? | Kayode |
| --- | --- |
| How does the newly added data compare with the original data? | Kayode |
| How was the data preprocessed? | Bhavyai |
| Preparing the model on Databricks with old + new data | Michael |
| How do the models perform on the original data vs the new + original data? | Michael |
| How does the performance of the models change based on the choice of hyper parameters? | Michael |
| How are the misclassifications of the best performing model distributed? | Bhavyai/Kayode |
| Discussion | All |

## 1.4 Databricks notebooks

1. README downloader
   https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/279226329024279/3713457635192600/4593657877666368/latest.html

2. Original Model Comparison
   a. ENSF612 Final Project with Original Models - Old Data: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4199039841175524/4424314020178626/1115343936761319/latest.html
   b. ENSF612 Final Project with Original Models – Old + New Data: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4199039841175524/4424314020178662/1115343936761319/latest.html

3. New Model Comparison:
   a. ENSF612 Final Project with Unexplored Models - Old Data: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4199039841175524/4424314020178557/1115343936761319/latest.html
   b. Histogram Gradient Boost – Old Data: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4199039841175524/4424314020178505/1115343936761319/latest.html
   c. Histogram Gradient Boost – Old+New Data: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4199039841175524/4424314020178533/1115343936761319/latest.html

4. Hyperparameter Optimization
   https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/279226329024279/105533218668830/4593657877666368/latest.html

# 1 ABSTRACT

## 1.1 Context

README files play an essential role in shaping a developer's first impression of a software repository and in documenting the software project that the repository hosts. Yet, we lack a systematic understanding of the content of a typical README file as well as tools that can process these files automatically.

## 1.2 Objective

In this project, we study how the understanding of content of a typical README can be improvised by categorizing the sections of README files using a multi-label classifier.

## 1.3 Methods

We replicate the research paper on "Categorizing the Content of GitHub README Files" [1]. We also extend the work done in the paper by running more ML classifier models. We also tune hyper-parameters of a few models to explore better performance.

## 1.4 Results

We find that the LinearSVC classifier used by the authors in their research paper is still the best performing model on default hyperparameters. The weighted F1 score of the LinearSVC we achieved is 0.721.

## 1.5 Conclusions

Through this work, we enable the owners of software repositories on sites such as GitHub to improve the quality of their documentation, and to make it easier for the users of the software held in these repositories to find the information they need.

# 2 INTRODUCTION

The README.md file for a repository on GitHub is often the first project document that a developer will see when they encounter a new project. This first impression is crucial.

With more than 25 million active repositories at the end of 20171, GitHub is the most popular version control repository and Internet hosting service for software projects. When setting up a new repository, GitHub prompts its users to initialize the repository with a README.md file which by default only contains the name of the repository and is displayed prominently on the homepage of the repository.

However, up to now and apart from some anecdotal data, little is known about the content of these README files.

To close this gap, the authors of this project manually annotated 1000 sections belonging to 86 README files. This manually annotated data is used in adding to the 4,226 annotations done by the authors of the research paper. This annotation provides the extended large scale empirical data on the content of GitHub README files.

In addition to the annotation, we use the classifiers used in the research paper and a set of features to predict categories of sections in the README files. We extend the work in the research paper by exploring more classifiers and tuning hyperparameters of some of the classifiers.

**Background**

GitHub is a code hosting platform for version control and collaboration.4 Project artifacts on GitHub are hosted in repositories which can have many branches and are contributed to via commits. Issues and pull requests are the primary artifacts through which development work is managed and reviewed.

Due to GitHub's pricing model which regulates that public projects are always free, GitHub has become the largest open source community in the world, hosting projects from hobby developers as well as organizations such as Adobe, Twitter, and Microsoft.

Each repository on GitHub can have a README file to "tell other people why your project is useful, what they can do with your project, and how they can use it." README files on GitHub are written in GitHub Flavored Markdown, which offers special formatting for headers, emphasis, lists, images, links, and source code, among others. In 2017, 25 million active repositories were competing for developers' attention, and README files are among the first documents that a developer sees when encountering a new repository.

## 3  RESULTS

### 3.1  How was the new data labeled/collected?

The new data was collected by developing a script which downloads README.md files from GitHub using GitHub API. The databricks version of this script is linked in the summary section 1.4.1. The downloaded files were randomly chosen and unique. We ensured that the files that were downloaded belong only to the software development repository and have size that is greater than 2KB. Manual filtering was done to remove non-English readme files. There is a limit to the number of readme files can be download from GitHub. The default maximum is 60 request per hour. In order the download more GitHub files for the project, we made use of Personal Access Token (PAT). This allowed us access to download up to 5000 request per hour. With this we were able to collect enough README files for the project.

Manual annotation was carried out on 1089 sections of the new dataset. We followed the same method used by the original authors to manually label each section of the README file in eight different categories. Each person in the team manual carried out the labelling separately. The results of the manual labelling were fed into IBM's SPSS software to compute Cohen Kappa inter-rater agreement metric. The result of the computation returns an agreement of 0.941 as shown below.

Though we had a very good metric in terms agreement, we still went ahead to carefully analyzed the area of disagreement and we were able to select the best categorization for each section where we disagreed based on majority opinion.

**Symmetric Measures**

| | | Value | Asymptotic Standard Error [a] | Approximate T [b] | Approximate Significance |
|---|---|---|---|---|---|
| Measure of Agreement | Kappa | .941 | .009 | 56.889 | .000 |
| N of Valid Cases | | 1089 | | | |

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 1: Summary of disagreement opinion statistics

## 3.2  How does the newly added data compare with the original data?

The new set data followed the same pattern as the original data. The table below shows the distribution when the old data and the new data were compared.

| Section Type | Original files count | % Count | New file sections | % count |
|---|---|---|---|---|
| What | 707 | 14.67% | 216 | 20.21% |
| Why | 116 | 2.41% | 0 | 0.00% |
| How | 2467 | 51.18% | 540 | 50.51% |
| When | 180 | 3.73% | 30 | 2.81% |
| Who | 322 | 6.68% | 73 | 6.83% |
| References | 858 | 17.80% | 179 | 16.74% |
| Contributions | 112 | 2.32% | 26 | 2.43% |
| Other | 58 | 1.20% | 5 | 0.47% |

Table 1: Distribution of old data and new data

We can clearly observe that sections on "How" have the highest percentage in both set of that. This is not surprising as developers tends to spend more time on explaining how to; run, install, update, set up download and fix errors. This explains why we have around 50% counts for the "how" section in each set of data. In contrary, the sections that fall under "other" have the least count as indicated by the table. The reason for this is that a README file section will only be categorized as "other" if there are no matching keywords used in determining the other categories.

## 3.3  How was the data preprocessed?

There are two pre-processing performed on the data. The headings and contents of the readme sections are abstracted to their types. Then this abstracted data is tokenized and stop words are removed.

Content abstraction abstracts contents to their types. We abstract the following types of section content: hyperlink, code block, image, and numbers. Each type is abstracted into a different string (@abstr_hyperlink, @abstr_code_section, @abstr_image and @abstr_number, respectively). Such abstraction is performed since for classification, we are more interested in existence of those types in a section than its actual content.

| Type | Abstracted text |
| --- | --- |
| Code snippets | @abstr_code_section |
| Numbers | @abstr_number |
| Images | @abstr_image |
| Hyperlinks | @abstr_hyperlink |
| mailto links | @abstr_hyperlink |

Figure 2: Summarizing the abstraction types

This abstraction is followed by tokenization, which converts a section into its constituent words, and English stop word removal. For the stop word removal, we use the stop words provided by scikit-learn.

We also encode our target readme section 1 to 8 into a matrix of 0s and 1s using MultiLabelBinarizer. The output of the MultiLabelBinarizer looks like Figure 2.

```
Section categories
['-', '1', '3', '4', '5', '6', '7', '8']

Encoded section categories in the sample readme file
[[0 0 1 0 0 0 0 0]
 [0 0 0 0 0 0 0 1]
 [0 1 1 0 0 0 0 0]
 [0 0 1 0 0 0 0 0]
 [0 0 1 0 0 0 0 0]
 [0 0 0 0 0 0 0 1]]
```

Figure 3: Sample output of a MultiLabelBinarizer

After pre-processing, we extract features from the data. We extract two kind of features – statistical features and heuristic features.

For statistical features, we count the number of times a word appears in each section. This is called the Term Frequency (TF) of a word in a section. If there are n words that appear in the set of sections used for training the classifier (after preprocessing), we would have n statistical features for each section. If a word does not appear in a section, then its TF is zero. We also compute the Inverse Document Frequency (IDF) of a word. IDF of a word is defined as the reciprocal of the number of sections in which the word appears. We use a multiplication of TF and IDF as an information retrieval feature for a particular word. A TF-IDF matrix from a sample README file is shown in Figure 3.

```
    abstr_code_section  abstr_hyperlink  abstr_image  abstr_number  ...        use     using       vcs  version
0             0.000000         0.326924     0.000000      0.490387  ...   0.000000  0.000000  0.000000  0.23611
1             0.000000         0.000000     0.000000      0.748083  ...   0.000000  0.000000  0.221518  0.00000
2             0.000000         0.000000     0.000000      0.000000  ...   0.000000  0.000000  0.000000  0.00000
3             0.000000         0.328074     0.473882      0.328074  ...   0.000000  0.000000  0.155436  0.00000
4             0.704681         0.000000     0.000000      0.000000  ...   0.140936  0.281872  0.000000  0.00000
5             0.000000         0.860429     0.000000      0.000000  ...   0.000000  0.000000  0.000000  0.00000

[6 rows x 54 columns]
```

Figure 4: TF-IDF matrix of a sample readme file

For heuristic features, we replicate the functions used in the research paper that extract 55 binary linguistic patterns within a category of sentences to derive heuristics that can aid classification. These heuristic patterns are categorized into 4 types, namely – Linguistic patterns, Single word non English heading, repository name, and non-ascii content text. A screenshot that captures some of the heuristic features extracted from a sample readme file is below
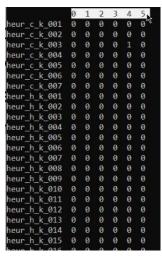
```
              0  1  2  3  4  5
heur_c_k_001  0  0  0  0  0  0
heur_c_k_002  0  0  0  0  0  0
heur_c_k_003  0  0  0  0  1  0
heur_c_k_004  0  0  0  0  0  0
heur_c_k_005  0  0  0  0  0  0
heur_c_k_006  0  0  0  0  0  0
heur_c_k_007  0  0  0  0  0  0
heur_h_k_001  0  0  0  0  0  0
heur_h_k_002  0  0  0  0  0  0
heur_h_k_003  0  0  0  0  0  0
heur_h_k_004  0  0  0  0  0  0
heur_h_k_005  0  0  0  0  0  0
heur_h_k_006  0  0  0  0  0  0
heur_h_k_007  0  0  0  0  0  0
heur_h_k_008  0  0  0  0  0  0
heur_h_k_009  0  0  0  0  0  0
heur_h_k_010  0  0  0  0  0  0
heur_h_k_011  0  0  0  0  0  0
heur_h_k_012  0  0  0  0  0  0
heur_h_k_013  0  0  0  0  0  0
heur_h_k_014  0  0  0  0  0  0
heur_h_k_015  0  0  0  0  0  0
heur_h_k_016  0  0  0  0  0  0
```

Figure 5: Part of 55 heuristic features of a sample readme file

### 3.4 How do the models perform on the original data vs the new + original data?

With the original data, we have performed model analysis of the weighted average of precision, recall and f1 score for the following models:

| Model Name | Weighted Avg | | |
| --- | --- | --- | --- |
| | Precision | Recall | f1 |
| RandomForestClassifier | 0.773 | 0.684 | 0.707 |
| LinearSVC | 0.709 | 0.738 | 0.721 |
| GaussianNB | 0.402 | 0.662 | 0.492 |
| LogisticRegression | 0.593 | 0.607 | 0.566 |
| | | | |
| BaggingClassifier | 0.634 | 0.731 | 0.673 |
| ExtraTreesClassifier | 0.813 | 0.636 | 0.692 |
| DecisionTreeClassifier | 0.619 | 0.704 | 0.652 |
| AdaBoostClassifier | 0.558 | 0.801 | 0.655 |
| HistGradientBoostingClassifier | 0.663 | 0.779 | 0.711 |

Table 2: Summarizing scores on all the models on original data

The first four models, namely RandomForestClassifier, LinearSVC, GaussianNB and LogisticRegression were used in the original paper.

The last five models, BaggingClassifier, ExtraTreesClassifier, DecisionTreeClassifier, AdaBoostClassifier and HistGradientBoostingClassifier are models that have not been experimented by the original paper. Note that initially the original GradientBoost model is used but Databricks terminated the cluster before the model can finish running. HistGradientBoostingClassifier is also significantly longer to complete than the other models (took 5 hours compared to 2.35hrs for RandomForest and 11.5 minutes for LinearSVC), but it can successfully complete with the scores calculated.

As can be seen from the previous table, ExtraTrees classifier has the highest precision whereas AdaBoostClassifier has the highest Recall.

Recall that Precision = (True Positive/Total Predicted Positive) and Recall = (True Positive)/(Total Actual Positive), Extra Trees Classifier is the best model to use when the cost of any false positive is high and AdaBoost should be used if we want to filter out the most false negatives.

Now, for overall performance we should instead look at what we call the f1 score, which is a function of Precision and Recall where the formula is:

F1 = 2*(Precision*Recall)/(Precision + Recall)

From the table, the top 3 models that have the highest weighted average f1 score are: LinearSVC (f1 score = 0.721), RandomForestClassifier (f1 score = 0.707) and HistGradientBoost (f1 score = 0.711).

For the new combined data, due to additional processing time required for the addition of the new dataset, we only perform validation on the top 3 models with the highest f1 score:

| Model Name | Precision Old | Precision New | Recall Old | Recall New | f1 Old | f1 New |
|---|---|---|---|---|---|---|
| | | | Weighted Avg | | | |
| LinearSVC | 0.709 | 0.7 | 0.738 | 0.74 | 0.721 | 0.716 |
| RandomForestClassifier | 0.773 | 0.763 | 0.684 | 0.68 | 0.707 | 0.7 |
| HistGradientBoost | 0.663 | 0.653 | 0.779 | 0.781 | 0.711 | 0.706 |

Figure 6: Summarizing old and new scores of top three models

As can be seen, the Precision, Recall and f1 scores drop slightly when additional data are added, but they do not affect the order of the performance ranking of these models.

### 3.5 How does the performance of the models change based on the choice of hyperparameters?

Attempts have been made to explore the hyperparameters for the project. We first explore the performance of the LinearSVC model based on the choice of hyper parameters. With LinearSVC, we explored the regularization parameter,

or the C parameter of the model. We will explore the effects on the model when parameter C is set to 0.001, 1, 100 and 10000.

For the purpose of conducting gridsearch, we developed a new method perform_grid_search as shown in Figure 7.

Grid Search to tune HyperParameters

```
def perform_grid_search(model, parameter_grid):
    '''
    Performs the grid search
    Prints the summary of of best model
    Returns the grid search object
    '''

    grid_search = GridSearchCV(model, parameter_grid, n_jobs=4, cv=10, scoring='average_precision', return_train_score=True)
    grid_search.fit(features_combined.values, labels_matrix)

    print("Best parameters    = {}".format(grid_search.best_params_))
    print("CV Training Score = {:.3f}".format(grid_search.cv_results_['mean_train_score'].max()))
    print("CV Test Score     = {:.3f}".format(grid_search.cv_results_['mean_test_score'].max()))

    return grid_search
```

Figure 7: Screenshot of grid search function

However, it seems like only the default C value (1) can be successfully completed as the other C values fails to converge:

```
  warnings.warn("Liblinear failed to converge, increase "
/databricks/python/lib/python3.8/site-packages/sklearn/svm/_base.py:985: ConvergenceWarning: Liblinear
ions.
  warnings.warn("Liblinear failed to converge, increase "
/databricks/python/lib/python3.8/site-packages/sklearn/svm/_base.py:985: ConvergenceWarning: Liblinear
ions.
  warnings.warn("Liblinear failed to converge, increase "
/databricks/python/lib/python3.8/site-packages/sklearn/svm/_base.py:985: ConvergenceWarning: Liblinear
ions.
  warnings.warn("Liblinear failed to converge, increase "
Best parameters    = {'estimator__C': 1, 'estimator__max_iter': 5000}
CV Training Score = 0.987
CV Test Score     = 0.657
```

Figure 8: Screenshot of convergence warnings on LinearSVC

We have also explored GridSearch using the Extra Trees, Ada Boost, and RandomForest and have also found that the default parameters were having the best results in most cases.

### 3.6  How are the misclassifications of the best performing model distributed?

The LinearSVC produced a total on 558 misclassifications on the test data of size 1060. We randomly pick 200 misclassifications and try to gauge which points might have led to misclassification, for example, presence of some keywords, links, the GitHub project name, etc. An excel containing the reasons of 200 misclassifications is being shared along with this report.

Some of the key findings labelling the misclassifications are listed below –
> (a) Sections often get labelled as 1 (Introduction) whenever explanation is given for some feature, process, etc.

(b)  Presence of links can potentially bias towards section getting classified as 6 (References)

(c)  Section 3 (How) get labelled as Section 6 (References) as their differences can sometimes be narrow

## 4  DISCUSSIONS

### 4.1  Bhavyai's discussion on the usage of models to write better README's

A real-life application of this model would be to generate badges based on the ML model output for each section. The generated badges can then be appended to every section heading. These badges help in increasing the readability of the readme file because they provide information at a glance about what each section talks about in the README file. They can also help visitors or other software developers in getting to familiarize with the repository quickly and easily.

Another application of this model would be to find what sections are missing from the README file. This could help the author of the repository to include all relevant information and write better READMEs.

### 4.2  Kayode's discussion

One way I can imagine is using models identify some ambiguous keywords in the README files sections. Once those ambiguities were identified and remove, it will be easy for the developers to pass across their messages in clear and concise manner.

Also, it would be interesting to develop a recommender model that will guide the developers on how to arrange the section so that they logically follow each other sequentially. This will enable the reader to quickly move to any relevant section of interest.

### 4.3  Michael Lee's discussion on the implications of the developed models

One useful scenario that may be useful for real-life application of this model is that it can be used to gauge the effectiveness of the Readme files. For example, in the world of internet marketing, we gauge the effectiveness of copywriting by its ability to sale their products or what they call the conversion rate. Similarly, we can adopt this method by linking our model with statistics such as the number of stars in the repository or visitor stats to see what kind of information, if included in their readmes, will increase people to visit, branch or star the repositories, therefore giving information to developers as to what information they should include in their readme files to maximize the effectiveness and increase popularity of their repository.

## 5  CONCLUSIONS

A README file is often the first document that a user sees when they encounter a new software repository. README files are essential in shaping the first impression of a repository and in documenting a software project. Despite their

important role, we lack a systematic understanding of the content. of README files as well as tools that can automate the discovery of relevant information contained in them.

In this project, we have reported on a qualitative study which involved the manual annotation of 1,000 sections from 86 README files for repositories hosted on GitHub, in addition to the 4,226 sections manually annotated by the research paper. We identified eight different kinds of content and found that information regarding the 'What' and 'How' of a repository is common while information on the status of a project is rare. Our best performing classifier achieved an F1 score of 0.721.

Our findings provide a point of reference for repository owners against which they can model and evaluate their README files, ultimately leading to an improvement in the quality of software documentation. Our classifier will help automate these tasks and make it easier for users and owners of repositories to discover relevant information.

**REFERENCES**

1. G. A. A. Prana, C. Treude, F. Thung, T. Atapattu, and D. Lo, "Categorizing the Content of GitHub README Files - Empirical Software Engineering," SpringerLink, 12-Oct-2018. [Online]. Available: https://link.springer.com/article/10.1007/s10664-018-9660-3. [Accessed: 11-Dec-2021].
2. Gprana. "Gprana/READMEClassifier." GitHub, https://github.com/gprana/READMEClassifier.

**APPENDIX**

Appendix A: Misclassifications and Explanation

| file_id | section_id | local_readme_file | heading_markdown | Manual Classification | Model Classification | Reason of misclassification |
|---|---|---|---|---|---|---|
| 1 | 1 | 1gitGrey.dotfiles-2.md | # holman does dotfiles | 2 | 3 | Because the word "how" in the first sentence. |
| 1 | 6 | 1gitGrey.dotfiles-2.md | ## components | 6 | 3 | The model assumes instructions were being given |
| 1 | 8 | 1gitGrey.dotfiles-2.md | ## thanks | 5 | 3 | The model assumes "how" because of the active word "forked". |
| 2 | 1 | 2917.Readmore.js.md | # Readmore.js | 1 | 1,3 | Might be due to some words appearing to suggest instructions |
| 2 | 5 | 2917.Readmore.js.md | ### The callbacks: | 3 | 3,6 | The model intepretes some words to be instruction |
| 2 | 6 | 2917.Readmore.js.md | #### Callback example: | 3 | 3,6 | The model intepretes some words to be instruction |
| 2 | 9 | 2917.Readmore.js.md | ## CSS: | 1, 3 | 3 | The model probably missed the "what" part of the section |
| 4 | 1 | 3rinN3lson.ai-tree-plugin.md | # AI Tree Plugin | 1 | 1,3 | Probably focused on the word "tested" |
| 4 | 3 | 3rinN3lson.ai-tree-plugin.md | ## Files/Folders: | 6 | 3 | Intepreted the active words as instructions |
| 4 | 5 | 3rinN3lson.ai-tree-plugin.md | ## Examples | 3 | 3,6 | A short reference was made to a folder during the instruction process |
| 4 | 7 | 3rinN3lson.ai-tree-plugin.md | ### Flocking Example | 3 | 1,3 | A few line intepreted as introduction |
| 4 | 8 | 3rinN3lson.ai-tree-plugin.md | ## How to Include: | 3 | - | The model return none of te category. No words after the section heading. |
| 4 | 9 | 3rinN3lson.ai-tree-plugin.md | ### Loading the Plugin | 3 | 3,6 | Because of the presence of " src/folder" |
| 4 | 13 | 3rinN3lson.ai-tree-plugin.md | #### Denotation | 3 | 1,3 | Asumes instruction were also being given |
| 4 | 20 | 3rinN3lson.ai-tree-plugin.md | #### TimeTrial | 3 | 3,6 | Asumed some words relating to reference |
| 4 | 21 | 3rinN3lson.ai-tree-plugin.md | ### Outer Nodes | 3 | 1,3 | The first sentence contains discriptive word. |
| 4 | 25 | 3rinN3lson.ai-tree-plugin.md | ### Advanced Features | 3 | 1,6 | The first sentence contains discriptive word. |
| 4 | 28 | 3rinN3lson.ai-tree-plugin.md | #### Inverters | 1 | 3 | The first sentence contains discriptive word. |
| 4 | 29 | 3rinN3lson.ai-tree-plugin.md | #### Multi-Parenting | 3 | 1,3 | The first sentence contains discriptive word. |
| 4 | 30 | 3rinN3lson.ai-tree-plugin.md | ## Acknowledgements and Resources | 5 | 6 | Intepreted as a refrence by the model |
| 5 | 1 | 40a.hub.md | # git + hub = github | 1 | 3 | The model sees the command line instructions and classified as "how" |
| 5 | 5 | 40a.hub.md | #### Standalone | 3 | 1,3 | Assumed the instruction to be an introduction |
| 5 | 7 | 40a.hub.md | # Assuming `~/bin` is in your PATH: | 3 | - | Did not classify. Might be due to short phrase |
| 5 | 9 | 40a.hub.md | ### Shell tab-completion | 6 | 3 | Did not classify. Might be due to short phrase |
| 5 | 10 | 40a.hub.md | ## Meta | 6 | 5 | Did not classify. Might be due to short phrase |
| 6 | 1 | 4nakin.NewsBlur.md | # NewsBlur | 1 | 1,5,6 | Classified based name and twiiter account appearing |
| 6 | 3 | 4nakin.NewsBlur.md | ## Technologies | 1 | - | Model did not classify |
| 6 | 18 | 4nakin.NewsBlur.md | #### Database server | 3 | 1,3 | Probably because of short phrase |
| 6 | 19 | 4nakin.NewsBlur.md | #### Task server | 3 | 1,3 | Probably because of short phrase |
| 6 | 20 | 4nakin.NewsBlur.md | ## Keeping NewsBlur Running | 1, 3 | 3 | Model only focused on words indicating instructions |
| 6 | 22 | 4nakin.NewsBlur.md | ### Feedback | 3 | 1,5 | The model missed the explanation part of the section |
| 8 | 1 | 920496044.ArduGuitar.md | # The ArduGuitar | 1 | 1,4 | The model probably spotted the date and intepreted it as "version" |
| 8 | 2 | 920496044.ArduGuitar.md | ## | 8 | - | Model did not classify |
| 11 | 1 | abhisheksprakash.ServiceStack.md | ### Simple, Fast, Versatile and full-featured Services Framework | 1 | 1,5,6 | Probably spotted "Amazon" as name thereby classifying as "who". |

| 11 | 2 | abhisheksprakash.ServiceStack.md | ### [Generate Instant Typed APIs from within all Major IDEs!](https://github.com/ServiceStack/ServiceStack/wiki/Add-ServiceStack-Reference) | 1 | 3 | Assumed some of active words to be giving instructions |
|----|----|----|----|----|----|----|
| 11 | 3 | abhisheksprakash.ServiceStack.md | #### [VS.NET integration with ServiceStackVS](https://visualstudiogallery.msdn.microsoft.com/5bd40817-0986-444d-a77d-482e43a48da7) | 1 | 1,6 | Providing explanation of integrating with IDEs was taken as "reference" |
| 11 | 4 | abhisheksprakash.ServiceStack.md | #### [Xamarin Studio integration with ServiceStackXS](https://github.com/ServiceStack/ServiceStack/wiki/CSharp-Add-ServiceStack-Reference#xamarin-studio) | 1 | 1,6 | Providing explanation of integrating with IDEs was taken as "reference" |
| 11 | 5 | abhisheksprakash.ServiceStack.md | #### [Xcode integration with ServiceStackXC Plugin](https://github.com/ServiceStack/ServiceStack/wiki/Swift-Add-ServiceStack-Reference) | 1 | -,1 | Providing explanation of integrating with IDEs was taken as "reference" |
| 11 | 7 | abhisheksprakash.ServiceStack.md | #### [IntelliJ integration with ServiceStackIDEA](https://github.com/ServiceStack/ServiceStack/wiki/Java-Add-ServiceStack-Reference#install-servicestack-idea-from-the-plugin-repository) | 1 | 1,6 | Providing explanation of integrating with IDEs was taken as "reference" |
| 11 | 8 | abhisheksprakash.ServiceStack.md | #### [Eclipse integration with ServiceStackEclipse](https://github.com/ServiceStack/ServiceStack.Java/tree/master/src/ServiceStackEclipse#eclipse-integration-with-servicestack) | 1 | 3,6 | The model assumes the section is giving instructions |
| 11 | 12 | abhisheksprakash.ServiceStack.md | ### [Calling from Swift](https://github.com/ServiceStack/ServiceStack/wiki/Swift-Add-ServiceStack-Reference#jsonserviceclientswift) | 3 | -,3 | Got most classification right but couldn't understand some commad line argument |
| 11 | 13 | abhisheksprakash.ServiceStack.md | ### [Calling from Java](https://github.com/ServiceStack/ServiceStack/wiki/Java-Add-ServiceStack-Reference#jsonserviceclient-usage) | 3 | -,3 | Got most classification right but couldn't understand some commad line argument |
| 11 | 14 | abhisheksprakash.ServiceStack.md | ### [Calling from Kotlin](https://github.com/ServiceStack/ServiceStack/wiki/Kotlin-Add-ServiceStack-Reference#jsonserviceclient-usage) | 3 | -,3 | Got most classification right but couldn't understand some commad line argument |
| 11 | 15 | abhisheksprakash.ServiceStack.md | ### [Calling from jQuery using TypeScript](https://github.com/ServiceStack/ServiceStack/wiki/TypeScript-Add-ServiceStack-Reference) | 3 | -,3 | Got most classification right but couldn't understand some commad line argument |
| 11 | 16 | abhisheksprakash.ServiceStack.md | ### Calling from jQuery | 3 | -,3 | Got most classification right but couldn't understand some commad line argument |

| 11 | 19 | abhisheksprakash.ServiceStack.md | ### [Release Notes](https://github.com/ServiceStack/ServiceStack/blob/master/release-notes.md) | 6 | -,1,6 | The model assumes the section is giving background information for the project. |
|---|---|---|---|---|---|---|
| 11 | 21 | abhisheksprakash.ServiceStack.md | ### [Install ServiceStack via NuGet](https://servicestack.net/download). | 6 | 1,6 | The model assumes the section is giving background information for the project. |
| 11 | 22 | abhisheksprakash.ServiceStack.md | ### [Docs and Downloads for older v3 BSD releases](https://github.com/ServiceStackV3/ServiceStackV3) | 6 | -,1,6 | The model assumes the section is giving more information. |
| 11 | 23 | abhisheksprakash.ServiceStack.md | ### [Live Demos](https://github.com/ServiceStackApps/LiveDemos) | 6 | -,1,6 | The model assumes the section is giving background information for the project. |
| 11 | 26 | abhisheksprakash.ServiceStack.md | ## OSS Libraries used | 6 | 1,6 | The model assumes the section is giving background information for the project. |
| 11 | 27 | abhisheksprakash.ServiceStack.md | ## Similar open source projects | 5 | 1,6 | The model assumes the section is giving background information for the project. |
| 11 | 28 | abhisheksprakash.ServiceStack.md | ## Find out More | 6 | 1 | The model assumes the section is |
| 11 | 29 | abhisheksprakash.ServiceStack.md | ## | 8 | - | Model couldn't classify as "others" |
| 11 | 30 | abhisheksprakash.ServiceStack.md | ## Core Team | 5 | 6 | The model assumes the section is giving more infromation and related projects. |
| 13 | 1 | adamjerickson.gitbook.md | ## Getting started | 1, 2 | 3,6 | The model assumes section is giving installation instructions and is giving more information. |
| 13 | 2 | adamjerickson.gitbook.md | ## Usage examples | 1 | 3,6 | The model assumes the section is giving instructions on how to use GitBook. |
| 13 | 3 | adamjerickson.gitbook.md | ## Help and Support | 6 | 5,6 | The model assumes the section is talking about the people behind the project. |
| 13 | 5 | adamjerickson.gitbook.md | ## Publish your book | 3 | 3,6 | This model assumes the section is giving more information. |
| 14 | 1 | adepue.moto.md | # Moto - Mock Boto | 1 | - | The model ignored the section title |
| 14 | 2 | adepue.moto.md | # In a nutshell | 1 | 1,3 | Models assumes some active words as passing instructions. |
| 14 | 3 | adepue.moto.md | ### Another Example | 1 | 3 | Models assumes some active words as passing instructions. |
| 14 | 10 | adepue.moto.md | ## Thanks | 8 | 5 | Spotted the name "Gabriel Falcao" and classified as "who" |
| 15 | 1 | adover.frontend.md | ## We're hiring! | 1 | 5,6 | Classified based on "we" and the url. |
| 15 | 3 | adover.frontend.md | # Core Development Principles (lines in the sand) | 1 | 3 | Might have classified based on the word "apply". |
| 15 | 4 | adover.frontend.md | ## On the server | 3 | 3,6 | The model assumes the section is giving more information |
| 15 | 5 | adover.frontend.md | # New developers quick-start | 3 | 3,6 | The model assumes the section is providing multiple links that give more information. |
| 15 | 7 | adover.frontend.md | ### Automatic | 3 | 3,6 | The model assumes the section is providing more information. - "…information on how to do so is here…" |
| 15 | 8 | adover.frontend.md | ###Â Manual | 3 | -,3 | Classified the instruction correctly |
| 15 | 9 | adover.frontend.md | #### Configuration files | 3 | 3,6 | Asuumed the path to folder is "reference" |
| 15 | 10 | adover.frontend.md | #### [Homebrew](http://brew.sh/) | 3 | -,3 | Classified correctly, returned blank because of the command line instruction |
| 15 | 12 | adover.frontend.md | #### [Node.js](https://github.com/joyent/node/wiki/Installing-Node.js-via-package-manager) | 3 | -,3 | Classified correctly, returned blank because of the command line instruction |

| 15 | 13 | adover.frontend.md | #### Grunt (build tool) | 3 | -,3 | Classified correctly, returned blank because of the command line instruction |
| 15 | 16 | adover.frontend.md | #### [bundler](http://gembundler.com/) | 3 | -,3 | Classified correctly, returned blank because of the command line instruction |
| 15 | 18 | adover.frontend.md | #### [libpng](http://libpng.org/pub/png/libpng.html) | 3 | -,3 | Classified correctly, returned blank because of the command line instruction |
| 15 | 24 | adover.frontend.md | ## Troubleshooting | 3 | 3,6 | The model picked 6 after reading the "For more info…" sentence. |
| 15 | 27 | adover.frontend.md | ###phantomjs permissions errors (OSX) | 3 | -,6 | "ask IT" makes this section fall under "getting support" and hence References (6). Also model might have found non-ASCII characters (-) |
| 15 | 35 | adover.frontend.md | ###Nginx | 3 | 6 | "refer to" makes this section fall under "more information" and hence References (6) |
| 15 | 38 | adover.frontend.md | ## Useful information and hints | 3 | 6 | "Further information" make the model think that its References (6) |
| 15 | 41 | adover.frontend.md | ###Deploying | 3 | 1 | "Name of the library" could have caused to categorize as project background |
| 16 | 1 | adtools.clib2.md | # clib2 â€" An ISO 'C' (1994) compliant runtime library for AmigaOS | 1 | - | Non ASCII caused it to be exclusion |
| 16 | 3 | adtools.clib2.md | ## What does it do? | 2 | 1 | 'Why' combined into 'What' |
| 16 | 5 | adtools.clib2.md | ## Where does the source code come from? | 5 | 1 | Project background is also given |
| 16 | 6 | adtools.clib2.md | ## Limitations and caveats | 6 | 1 | Some background about documentation is given |
| 16 | 7 | adtools.clib2.md | ### Floating point math and functions (`scanf()`, `printf()`, etc.) | 3 | 3,6 | Linking to libc.a, and code samples might have made it to be classified also as "References" |
| 16 | 10 | adtools.clib2.md | ### Implementation defined behaviour | 3 | - | Might have incorrectly found non-ASCII characters |
| 16 | 11 | adtools.clib2.md | #### 'C' language | 3 | -,3 | Might have incorrectly found non-ASCII characters |
| 16 | 15 | adtools.clib2.md | #### Library functions | 3 | 1 | The explanations given could have made it look like background |
| 16 | 19 | adtools.clib2.md | ##### Signal handling | 3 | 3,6 | "library" word could have meant API |
| 16 | 32 | adtools.clib2.md | ## Conventions and design issues | 1 | 3 | The code snippets may have made it look like instructions |
| 16 | 35 | adtools.clib2.md | ## Legal status | 6 | 1,5 | The word BSD license and its explanation may have caused 1 and 5 |
| 16 | 36 | adtools.clib2.md | # <pre> | 8 | - | Non ASCII caused it to be exclusion |
| 16 | 37 | adtools.clib2.md | # is preserved. | 8 | -,3 | Non ASCII, action verb, and lack of more data |
| 17 | 1 | afeld.type-is.md | # type-is | 1 | 1,6 | Links may have caused for more information |
| 17 | 3 | afeld.type-is.md | ## API | 6 | 3 | Code snippet on how to run |
| 17 | 6 | afeld.type-is.md | ## Examples | 3 | 3,6 | Code snippet on how to run |
| 17 | 8 | afeld.type-is.md | ## License | 6 | 5 | The link to license |
| 18 | 7 | agiza.heartbeat.md | ## Resources/Providers | 3 | 6 | Inline code segments, "Providers" may have caused feedback |
| 18 | 14 | agiza.heartbeat.md | ## Recipes | 3 | 1 | The steps to install and explanation - caused it as introduction |
| 18 | 15 | agiza.heartbeat.md | ### default | 3 | 1,3 | The steps to install and explanation - caused it as introduction |
| 19 | 1 | agrede.SAMIS.md | # SAMIS | 1 | 6 | The links would have made it like more information |
| 20 | 2 | Ahmad-me.osmbonuspack.md | # Examples | 1 | 3 | Images make it like How to |
| 20 | 3 | Ahmad-me.osmbonuspack.md | ## on the map \| as a list view | 1 | - | Non ASCII |

| 20 | 4 | Ahmad-me.osmbonuspack.md | ## Google Maps "My Places" \| rendered with OSMBonusPack | 1 | 3 | Images make it like How to |
|---|---|---|---|---|---|---|
| 20 | 5 | Ahmad-me.osmbonuspack.md | # How to use it | 3 | 6 | Links may have caused for more information |
| 20 | 6 | Ahmad-me.osmbonuspack.md | # How to get help | 3 | 6 | "Request" and Stackoverflow may have caused it as more information |
| 21 | 1 | ahmedbodi.solarcoin.md | # SolarCoin integration/staging tree | 1 | 5 | Image and copyright info could have caused it to under section 5 |
| 21 | 2 | ahmedbodi.solarcoin.md | ## What is SolarCoin? | 1 | 1,4 | Mentioning of binary version caused it to go udner When |
| 21 | 3 | ahmedbodi.solarcoin.md | ## License | 6 | 5 | Mention of MIT may have caused under Who |
| 21 | 4 | ahmedbodi.solarcoin.md | ## Development process | 1 | 4,7 | The branch names may have caused it to ne under When and Contribution |
| 21 | 5 | ahmedbodi.solarcoin.md | ## Testing | 3 | 1,6,7 | Mention of pull requests and explain of testing caused it to be under |
| 22 | 3 | ahmedkato.countly-server.md | ##Supported devices | 3 | 1,3 | Explanation of supported devices, and links may have caused as Introduction as well |
| 22 | 5 | ahmedkato.countly-server.md | ##Dependencies | 3 | 1,3 | Description of dependencies caused it as Introduction as well |
| 22 | 6 | ahmedkato.countly-server.md | ##API & Frontend | 6 | 3,6 | Examples of inline code segments probably caused it to label as "How to" |
| 22 | 7 | ahmedkato.countly-server.md | ##How can I help you with your efforts? | 3 | 3,7 | Mention of pull requests caused it under References |
| 22 | 8 | ahmedkato.countly-server.md | ##Links | 6 | 1,6 | Links, and explanation of enterprise edition caused it as Introduction\ |
| 24 | 1 | airdesigns1.mobile-samples.md | # Mobile Samples | 1 | 3 | "How to" string cause it as "How" |
| 24 | 2 | airdesigns1.mobile-samples.md | ## License | 6 | 5 | Mention of word License caused it under Who |
| 24 | 4 | airdesigns1.mobile-samples.md | ## Samples Contribution Guidelines | 3 | 5,7 | Word contribution probably has caused it under Who and Contribution |
| 24 | 7 | airdesigns1.mobile-samples.md | ## Sample Requirements | 3 | 3,6 | Welcome sample submissions may have caused it under References |
| 24 | 8 | airdesigns1.mobile-samples.md | ## GitHub Integration | 3 | 3,6 | Instructions on committing may have caused it to get labelled as References |
| 25 | 1 | ajcrowe.pubsubbeat.md | # Pubsubbeat | 1 | 3,6 | The presence of inline code, and word github.com could have caused it as How and References |
| 26 | 1 | ajiangcn.Probabilistic-Programming-and-Bayesian-Methods-for-Hackers.md | #[Bayesian Methods for Hackers](http://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/) | 1 | 1,6 | Links might have made it to be classified also as "References" |
| 26 | 4 | ajiangcn.Probabilistic-Programming-and-Bayesian-Methods-for-Hackers.md | ## Contents | 6 | 1 | Links, and explanation of sections caused it as Introduction |
| 26 | 5 | ajiangcn.Probabilistic-Programming-and-Bayesian-Methods-for-Hackers.md | ## Using the book | 3 | 1,3 | Explanation of methods, and links may have caused as Introduction as well |
| 26 | 11 | ajiangcn.Probabilistic-Programming-and-Bayesian-Methods-for-Hackers.md | ## Reviews | 5 | 1,6 | The review explains about the book, caused it as Introduction and author names caused it as References |
| 27 | 1 | ajita.jnode.md | # jnode | 1 | 1,3 | Mention of functions caused it under How to |
| 27 | 2 | ajita.jnode.md | # How to get jnode | 3 | -,3 | Non ASCII could have caused it to be exclusion |
| 27 | 3 | ajita.jnode.md | # How to run jnode | 3 | -,3 | Non ASCII could have caused it to be exclusion |
| 27 | 7 | ajita.jnode.md | # Query String | 3 | - | Non ASCII could have caused it to be exclusion |

| 28 | 1 | akerekes.redis-collectd-plugin.md | # redis-collectd-plugin | 1 | 1,3 | Listing of inclusions of data captured, and links might have caused it as How |
|---|---|---|---|---|---|---|
| 29 | 2 | AKMCC-hub.codeigniter-oauth2-server.md | ## | 8 | - | Absence of text caused it as exclusion, while we categorized it as Other |
| 29 | 7 | AKMCC-hub.codeigniter-oauth2-server.md | # Features | 3 | 1 | Support listing may have caused it as Introduction |
| 29 | 9 | AKMCC-hub.codeigniter-oauth2-server.md | # License | 6 | 5 | Mention of MIT may have caused under Who |
| 31 | 1 | akshay12489.StupidPetTricks_I.md | # StupidPetTricks I | 3 | - | The model couldn't classify it and put under exclusion |
| 31 | 2 | akshay12489.StupidPetTricks_I.md | #IDM 9013A: Programming Physical Objects | 3 | 8 | The model couldn't classify it and put under exclusion. May be less info caused it to get misclassified |
| 31 | 3 | akshay12489.StupidPetTricks_I.md | ##Stupid Pet Tricks | 3 | -,6 | The model couldn't classify it and put under exclusion. May be less info caused it to get misclassified |
| 32 | 1 | akshaydashrath.sms-backup-plus.md | ## SMS Backup+ | 3 | 1 | Some background about How to is given |
| **32** | **5** | **akshaydashrath.sms-backup-plus.md** | **### Initial backup** | **3** | **1** | **Some background about working of the backups is given** |
| 32 | 7 | akshaydashrath.sms-backup-plus.md | ### <a name="call-log-support">Call log support</a> | 3 | 1,3 | Some background about working of the logs is given |
| 32 | 8 | akshaydashrath.sms-backup-plus.md | ### <a name="whatsapp-support">WhatsApp support</a> | 3 | 1,6 | Some background about working of the Whatsapp support is given |
| 32 | 9 | akshaydashrath.sms-backup-plus.md | ### <a name="3rdparty">3rd party app integration</a> | 3 | 6 | "3rd party" and send could have caused it as References |
| 32 | 12 | akshaydashrath.sms-backup-plus.md | ### General questions | 3 | 6 | The faqs made it look like more information and getting support |
| 32 | 13 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-permissions">Why does it need so many permissions?</a> | 3 | 1 | Some background about permissions is given |
| 32 | 16 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-can-you-add-feature-x">Can you add feature X ?</a> | 3 | 1 | Some background about cluttering of features is given |
| 32 | 17 | akshaydashrath.sms-backup-plus.md | ### Backup questions | 3 | 6 | Probably the images, important, and Gmail labelled it as References |
| 32 | 18 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-inbox">Why do backed up SMS show up in my inbox?</a> | 3 | 1 | Some background about cluttering of features is given |
| 32 | 21 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-reset">How can I make the app think that it has to do the backup again?</a> | 3 | 1 | Some background about reset is given |
| 32 | 22 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-schedule">What's the difference between regular and incoming backup schedule?</a> | 3 | 1 | Some background about backup schedules is given |
| 32 | 23 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-scheduling">I'd like SMS Backup+ to schedule a backup only at a given time of the day / when Wifi is available / etc.</a> | 3 | 1 | Some background about backup schedules is given |
| 32 | 28 | akshaydashrath.sms-backup-plus.md | #### <a name="whatsapp-does-not-get-backed-up">I enabled the WhatsApp backup, but my messages don't get backed up!</a> | 3 | 1,6 | Some background about whatsapp backup and the linkschedules is given which got it labelled as Introduction and References |
| 32 | 29 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-untrusted-certificate">I get the error "Trust anchor for certification path not found"</a> | 3 | 1,3 | Some background about certificates and connection is given |

| 32 | 30 | akshaydashrath.sms-backup-plus.md | #### <a name="faq-default-app">Why does SMS Backup+ ask to become the default SMS app?</a> | 3 | 1 | Some background about default app is given |
|----|----|-----|-----|-----|-----|-----|
| 32 | 33 | akshaydashrath.sms-backup-plus.md | ### Authentication questions | 3 | 6 | The word "questions" caused it as References |
| 32 | 36 | akshaydashrath.sms-backup-plus.md | ### Device specific questions | 3 | 6 | The word "questions" caused it as References |
| 32 | 38 | akshaydashrath.sms-backup-plus.md | ## <a name="beta">Beta testing | 3 | 6 | The section mentions the G+ community which makes it fall under References |
| 32 | 39 | akshaydashrath.sms-backup-plus.md | ## <a name="contributing">Contributing</a> | 3 | 7 | The word contribution caused it under Contribution |
| 32 | 41 | akshaydashrath.sms-backup-plus.md | ### <a name="translating">Translating the UI</a> | 3 | 3,6 | Links, pull request might have caused to it labelled as References |
| 32 | 42 | akshaydashrath.sms-backup-plus.md | ##<a name="credits">Credits</a> | 3 | 5,6 | The listing of names of people, and word Credit could have caused it as Who |
| 32 | 43 | akshaydashrath.sms-backup-plus.md | ##<a name="screenhots">Screenshots</a> | 3 | 3,8 | Image and lack of textual info could have caused it to be under Other |
| 32 | 44 | akshaydashrath.sms-backup-plus.md | ##<a name="license">License</a> | 3 | 5 | Mention of "License" caused it to be under Who |
| 33 | 1 | Alan0147.android-samples.md | # Google Maps Android API v2 Samples | 1 | 1,3,6 | Links, and explanation of components, word "Tutorial" caused labelling as 3 and 6 |
| 33 | 2 | Alan0147.android-samples.md | ## Pre-requisites | 1 | 3 | The word requisites caused it to be under How |
| 33 | 3 | Alan0147.android-samples.md | ## Getting Started | 3 | 3,6 | Links, see options caused it as more information as hence References |
| 33 | 4 | Alan0147.android-samples.md | ## Support | 5 | 6 | Links, instructions caused it under References |
| 34 | 1 | alanszp.marvel.md | # angular-seed â€" the seed for AngularJS apps | 1 | 3 | Code snippet on how to run, starting up labelled it as How |
| 34 | 6 | alanszp.marvel.md | ### End to end testing | 3 | 1,3 | Some background about testing and usage is given |
| 34 | 8 | alanszp.marvel.md | ### Receiving updates from upstream | 3 | 1 | Some background about updates is given |
| 34 | 9 | alanszp.marvel.md | ## Directory Layout | 3 | 1,3 | Layout could have caused it as Background and hence What |
| 37 | 1 | Alex-Bubblemaster.ASP.NET-MVC.md | # Spec Tester Documentation | 8 | | Model couldn't classify so put null, we put 8 |
| 37 | 2 | Alex-Bubblemaster.ASP.NET-MVC.md | ## Home Page | 8 | 6 | "Home page" might have caused it to be under more information |
| 37 | 4 | Alex-Bubblemaster.ASP.NET-MVC.md | ## Admin Training Management | 8 | 1 | "Management" might have caused it to be under introduction |
| 37 | 5 | Alex-Bubblemaster.ASP.NET-MVC.md | ## Admin Products page - to Manage Products, create new etc. | 8 | 3,6 | "manage", "create new" could have caused it as API documentation and How to |
| 37 | 6 | Alex-Bubblemaster.ASP.NET-MVC.md | ## Admin Manage Area with Statistics | 4 | 1 | Enlisting of features may have caused it as project background. |
| 37 | 7 | Alex-Bubblemaster.ASP.NET-MVC.md | ## Cook book | 3 | 3,6 | "More" could have meant more information |
| 37 | 8 | Alex-Bubblemaster.ASP.NET-MVC.md | ## Error Handling | 3, 4 | 3,6 | "More" could have meant more information |
| 37 | 9 | Alex-Bubblemaster.ASP.NET-MVC.md | ## User statistics | 4 | -,3 | Classified wrongly probably because of "coming" word |
| 38 | 1 | AlexanderKovachev.SharpSapRfc.md | # SharpSapRfc | 8 | - | Could not classify the title. |
| 38 | 2 | AlexanderKovachev.SharpSapRfc.md | ## What is SharpSapRfc? | 1,3 | 3 | Assumed active words to be giving instructions |
| 43 | 1 | alirezarahmani.jquery-cookie.md | # jquery.cookie [![Build Status](https://travis-ci.org/carhartl/jquery-cookie.png?branch=master)](https://travis-ci.org/carhartl/jquery-cookie) [![Code Climate](https://codeclimate.com/github/carhartl/jquery-cookie.png)](https://codeclimate.com/github/carhartl/jquery-cookie) | 1,6 | 6 | Excluded the "what" part of the section. Therby clasified as "what". |

| 43 | 11 | alirezarahmani.jquery-cookie.md | ### domain | 3 | 3,6 | Asummed "example.com" as a reference. |
| 43 | 12 | alirezarahmani.jquery-cookie.md | ### secure | 3 | 6,8 | Misclassified wrongly. |
| 45 | 1 | alvarosperez.Mining-the-Social-Web-2nd-Edition.md | # Mining the Social Web (2nd Edition) | 8 | 1 | Asumedthe title to be giving instructions. |
| 45 | 2 | alvarosperez.Mining-the-Social-Web-2nd-Edition.md | ## Summary | 1,2,6 | 1,6 | Omiited the "why" part of the section |
| 45 | 3 | alvarosperez.Mining-the-Social-Web-2nd-Edition.md | ## Preview the Full-Text of Chapter 1 (Mining Twitter) | 6 | 1 | Classified based on some explanatory words e.g "provides" |
| 45 | 4 | alvarosperez.Mining-the-Social-Web-2nd-Edition.md | ## Preview the IPython Notebooks | 1,3 | 1,6 | Some of the instructions were taken as "references" |
| 45 | 9 | alvarosperez.Mining-the-Social-Web-2nd-Edition.md | ## "Premium Support" | 6 | 1,6 | The section mentioned the book name and the model classified it as refrence. |
| 46 | 1 | ALYREZA.docker-dns-ad-blocker.md | # oznu/dns-ad-blocker | 1 | -,1 | Got "what" right |
| 46 | 6 | ALYREZA.docker-dns-ad-blocker.md | # webserver. | 3 | | Did not classify the section based on a very shor phrase. |
| 46 | 8 | ALYREZA.docker-dns-ad-blocker.md | # servermachine.com and preference 50 | 3 | 6 | Asuumed the url to be a "refrence" |
| 48 | 1 | amykatenicho.IoTWorkshop.md | # Windows 10 IoT Core Hands-on Lab | 1,3 | 3 | Missed the "what" paer of the explanation |
| 48 | 5 | amykatenicho.IoTWorkshop.md | ### Download the IoT Core Dashboard | 3,6 | 3 | Intepreted tge reference to the link as part of the instructions |
| 48 | 8 | amykatenicho.IoTWorkshop.md | ### Setting up your Azure Account | 3 | 3,6 | Classified the the url as "reference". |
| 48 | 10 | amykatenicho.IoTWorkshop.md | ### Registering your device | 3 | 3,6 | Classified the the url as "reference". |
| 48 | 11 | amykatenicho.IoTWorkshop.md | ## Creating a Universal App | 3 | - | Missed the section probably because the title was not seperated |
| 48 | 13 | amykatenicho.IoTWorkshop.md | ## Summary | 6 | | Missed the section probably because the title was not seperated |
| 49 | 1 | anaderi.FairSoft.md | #FairSoft | 1,3,6 | 3,6 | Missed the "why" part probably because alll the categories were emebedded in the whole section. |
| 49 | 5 | anaderi.FairSoft.md | ##Included Packages | 3 | 4 | Might have read some version number as dates |
| 50 | 1 | analytically.slack-express.md | # slack-express [![bitHound Overalll Score](https://www.bithound.io/github/smallwins/slack-express/badges/score.svg)](https://www.bithound.io/github/smallwins/slack-express) [![bitHound Dependencies](https://www.bithound.io/github/smallwins/slack-express/badges/dependencies.svg)](https://www.bithound.io/github/smallwins/slack-express/master/dependencies/npm) [![Codeship Status for smallwins/slack-express](https://codeship.com/projects/8348a860-7a10-0133-8cf8-72bb2b768401/status?branch=master)](https://codeship.com/projects/118901) | 1 | 6 | Read the title as "references" |
| 50 | 4 | analytically.slack-express.md | ### middleware | 3 | 6 | Read some of the words e.g "slack.slash" as "references" |
| 50 | 6 | analytically.slack-express.md | ### persistence | 3 | 1,3 | Assumed the first sentence to be an instruction. |