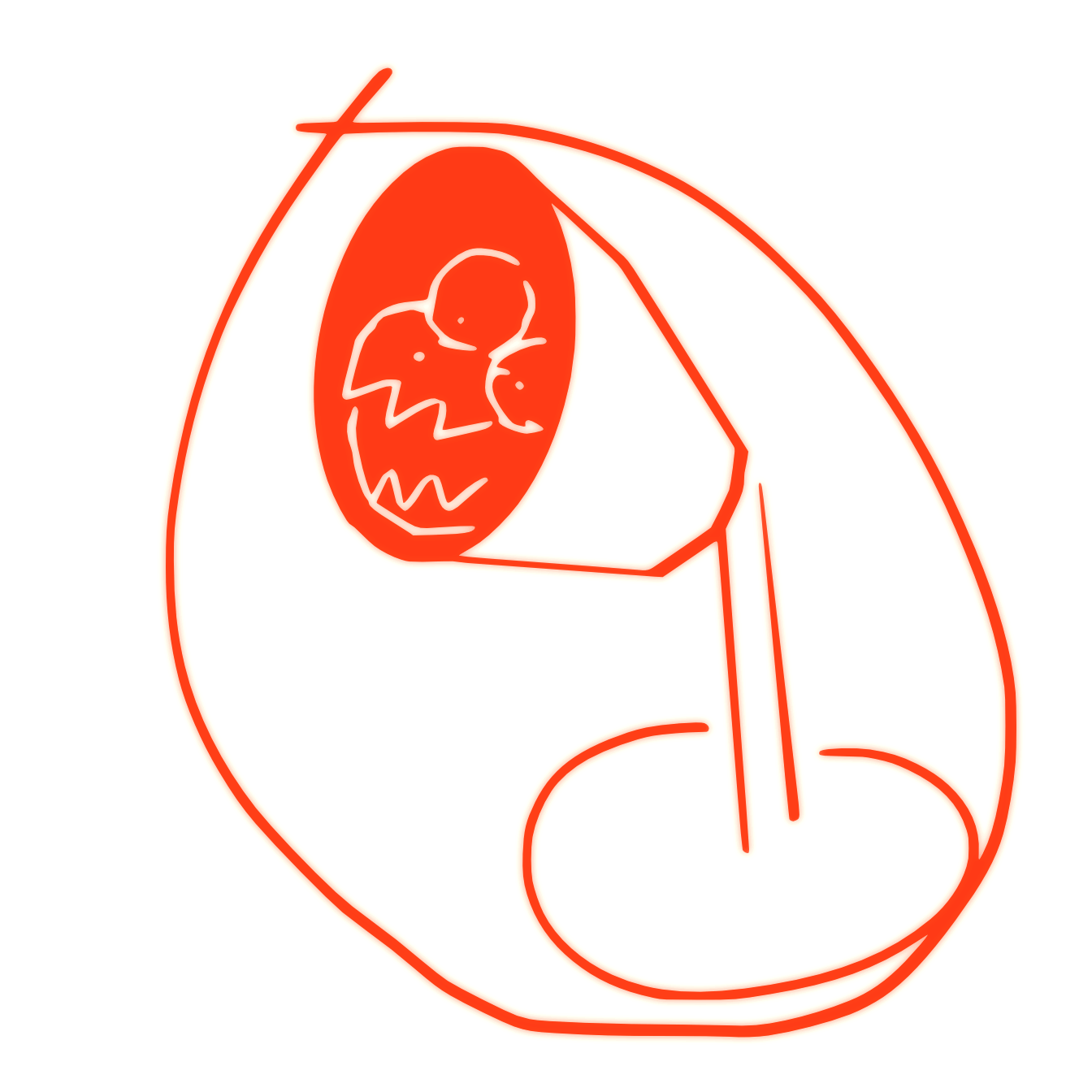
4/10/2013

|  |
| --- |
| [Type the document subtitle] | Akodiat |



|  |  |
| --- | --- |
| Chalmers | Sentimentalizer - LampMonster |

A good title for an interesting project

Group X: Author1, Author 2, Author 3, Author 4

XX May 2013

# Abstract

...no more than 10 lines!

Table of Contents

[Abstract 1](#_Toc353866549)

[1 Introduction and Background 2](#_Toc353866550)

[1.1 The problem you tried to solve 2](#_Toc353866551)

[1.2 Results from the literature 2](#_Toc353866552)

[1.3 What tools and programs are already available for the problem, or for closely related ones? 2](#_Toc353866553)

[2 Overview of the architecture 2](#_Toc353866554)

[2.1 Finished work: Running modules 2](#_Toc353866555)

[2.1.1 Naïve Bayes 2](#_Toc353866556)

[2.2 Work in progress: Modules designed but not implemented 2](#_Toc353866557)

[2.3 Future work: Modules a future continuation may have 2](#_Toc353866558)

[3 Results and Evaluation 2](#_Toc353866559)

[4 Discussion and Conclusions 2](#_Toc353866560)

[5 References 2](#_Toc353866561)

[Appendices 3](#_Toc353866562)

[A. Individual stories 3](#_Toc353866563)

[B, C, ... 3](#_Toc353866564)

# 1 Introduction and Background

A short general text about the area of ai that your project belongs to.

## 1.1 The problem you tried to solve

Here you describe the problem in your own words!

## 1.2 Results from the literature

Describe briefly the scientific papers or book chapters you found relevant to the problem, references to sec 6. Explain which are relevant for your project and which not and why.

## 1.3 What tools and programs are already available for the problem, or for closely related ones?

Describe these briefly. Say how you can use them, and how your work will build on them, or differ from them. Explain which are relevant for your project and which not and why.

# 2 Overview of the architecture

Describe the different parts of your program suite in detail.

## 2.1 Finished work: Running modules

### 2.1.1 Naïve Bayes

Naïve Bayes is based on Bayes theorem, making a lot of incorrect assumptions and still managing to perform well.

Basically, we want to calculate, that is, the probability of a specific class given a document. Bayes rule states that this is the same thing as . Once we have calculated for each class c, we choose the c with the responding highest probability, that is:

However is the same for all the classes, and doesn’t have to be included in the calculations since we are just comparing the relative sizes:

Since calculating , the probability of a document given a class is quite vague, hard to understand and even harder to calculate without simplification. So instead of doing this, we calculate the probability for all words in the document given a class:

And go on to simplify this even further by assuming that the order of the word doesn’t matter, and that the words are independent to each other. By making this assumptions we can calculate as . And thus:

…Where D is a bag of words representing the document to be classified.

As a further adjustment, log(x) could be, and is often, used instead of P(x). Instead, we have decided to still use P(x) and have a 128 floating point precision construct.

Basically, given a document and a category the algorithm will loop through all words in the document and for each of them calculate the probability that the word belongs to the given category. All these word probabilities are then multiplied together, and then multiplied with the probability of the category (normally 1/#ofCategories). All this will be done for all categories and the resulting values for the different categories will be compared. The category with the highest probability value will be the one the document is classified as.

As the probabilities of the words are multiplied together, the product would be zero if any of the probabilities are zero. Each probability is therefore calculated as where prior is a constant added as a Laplace smoothing. Without a prior, we would get zero as the probability product every time an unknown word (a word that wasn’t added to a category during training) appears in a document.

We use the value one as our prior, but it should be possible to change this value in order to improve performance. However, we decided to not make that our priority to test.

## 2.2 Work in progress: Modules designed but not implemented

## 2.3 Future work: Modules a future continuation may have

# 3 Results and Evaluation

What does your running code do? How does it fare against your benchmarks and instances? Describe advantages and disadvantages, possibly in relation to other groups in this course.

# 4 Discussion and Conclusions

Sum up your project, suggest future extensions and improvements.

# 5 References

|  |  |
| --- | --- |
| [1] | D. Jurafsky and C. Manning, "6 - 2 - Naive Bayes - Stanford NLP - Professor Dan Jurafsky & Chris Manning," 4 April 2012. [Online]. Available: https://www.youtube.com/watch?v=DdYSMwEWbd4. [Accessed 8 April 2013]. |

List relevant section numbers in the textbook, and full bibliographic references to relevant articles and papers you found (active web links help). This is a vital part of the document!

References should be cited (i.e. actually be referred to) at the appropriate place in your text, they should visibly influence your document, and they should convey as much information as possible to the reader.

A reference to a paper will start with a unique ID (numbers, letters, ...) and have the following three paragraphs:

1. Author(s) [year of publication], "Title of paper", Journal (or conference or publisher etc.), location data (i.e., volume numbers, etc. needed to ﬁnd the item), a live www link if available. Also, if the reference is to a specific topic in a book (or large survey article), it should include the section numbers or page numbers that deals with the material you are referring to.
2. A couple of lines saying what question the paper is attacking, and what the proposed solution is. Or what topic the reference deals with (your textbook has excellent examples in the sections entitled "Bibliographical and Historical notes").
3. A pointer back to where you cite this reference.

An example would look like this:

[REF1]

Turing, A.M. (1937). "On Computable Numbers, with an Application to the Entscheidungsproblem". Proceedings of the London Mathematical Society. Series 2, 42: 230–265.

The paper is about ...

The most important reference to this paper in our document is in Section XYZ

# Appendices

## A. Individual stories

Each group member should write one page where they describe their own contributions to the project.

## B, C, ...

Here you include all other information and documentation that did not fit into the report in the above sections but that you consider too important to leave out.