Note: (Report)

* I recommend using an Oxford comma.
* ‘However’, ‘Therefore’, ‘Thus’ are conjunctive adverbs. If liking two complete sentences together, a ‘semicolon’ rather than a ‘comma’ should precede conjunctive adverbs.
* Past tense should be used when referring to other people’s work.
* Table caption should be on top.
* In the method chapter, number indicating \subsection or \subsubsection would be good.
* Be consistent with the word using in the thesis. Ex. dataset or data set
* There should be no space between number and %
* data base 🡪 database
* In the results chapter, you used both present tense and past tense together when describing the obtained results. Please check and decide on the tense you want to use.
* For the misclassification rate of alpha, I think added dashed line for the optimal alpha would be nice.
* When describing the confusion matrix, I think it’s better to always include the number of class along with the name. The table only presents the number of class so it is not easy to follow.
* You used ‘Apriori rules or topics’ for the section. However, the caption of figure and table, and also some texts are written with ‘topics and Apriori’ instead. It’s not a big deal but I think it’s good to have consistent of the method name for the entire thesis. This applied to the SPADE as well.
* Again, I suggest you include the number of \subsection or \subsubsection for each model. Then, it would be easier when referring to this model (use cross-reference) and make it less confusing.
* It would also be nice to have a heading for the logistic regression and the DO-probit in each method. So, the reader will know that now you will be presented with another model.

4.5.1 Apriori

4.5.1.1 Logistic regression

4.5.1.2 DO-probit

4.5.2 Apriori and topics

4.5.2.1 Logistic regression

4.5.2.2 DO-probit

Q:

* Is ‘system log file’ and ‘alarm log file’ the same thing? If so, maybe you can state that as well to make it clearer.
* Why talk about DOLDA in the introduction? Is DO-probit an extension to DOLDA? Are there any relationship between these two method?
* Still not so clear about the process of bug report/ alarm log files and also description of data
* MO and alarm?
* What are 40 topics?
* Partially Collapsed Gibbs sampling 🡪 Why use these two parameters to collapse this parameter? Is it because of D
* Why do we need parallelizing? Computation cost?
* Topic indicators Z\_{1:D} 🡪 Does it mean topic indicator=topic assignments?
* Do you use ‘lift’ in the analysis? If not, I don’t think it is necessary to also mention it in the thesis.
* 183 🡪 184? Because the number of days in this range Jul-Dec, Aug-Jan, Mar-Jun are 184.
* The reason of choosing Beta distribution? Is it because of characteristic of the time elapse?
* Time dependent weight term: w\_{t} and all weights: lambda\_{ij}

lamba\_{ij} = w\_{t}?

The last occurring of the alarm has highest value? max(weight)

* A time difference weight: w\_{ir}
* What is the pattern r?
* MO and alarms? What is alarms?
* Are the variables in the data highly correlated?