



# **News Aggregator with Python and Machine Learning 1st Audit**

Client: Thomas Griffith  
Team Members: Yuliang Zhang, Jiahua Liang, Chen Zhang, Luokun Gong, Vishnu Vardhan Jasti, Jun Yang

# Background



## News Channel Problem:

- Traditional Media Recession
- New Media Fragmentisation & Vulgar

Difficult to get **high-value** news **efficiently**



## Solution:

- A News Aggregator Focusing on **High-Value** News
- Using Web Crawler and Machine Learning Tools

<http://gpu.jkl.io:5000/>



## Continuous Project Problem:

- Vague User Scope
- Rely on Algorithm, difficult to estimate user satisfaction
- Insufficient news channels
- Existing Code Bugs
  - Bad BERT Results
  - Crawler Stops Occasionally
  - Only 1 news source could be stored in data base

Solution??

# Client's Expectation & Requirements



## More Canberra Focused:

- More Extraction Work on Canberra Based News Channels
- Improve Algorithm for Canberra residents



## User Surveys:

- Conduct 2-3 User Surveys
- Improve our work according to user feedback



## Improve Web UI

# Stakeholders

Client: Thomas Griffith



Provide: development resources  
•server, web frame, data, etc  
Expect:  
•A Canberra Focused, operable news website  
•2-3 User Survey Reports

Team Members



Provide:  
•Time and effort  
Expect:  
•Improve code skills  
•Improve teamwork skills  
•Get good marks

Users



Provide:  
•Time to use the website  
•Time to do user survey  
Expect:  
•A efficient website for getting high-value news

## Resources & Costs



### Resources:

- Kosmos Dataset (news dataset including more than 300,000 articles)
- GPU servers (used for training machine learning model)
- The code source and products from last semester's work
- The Bootstrap or other frames for web development
- Some Canberra focused news websites.



### Costs:

Development related costs will be solved by our client.

# Risks



## Risks:

- Fail to achieve the expected accuracy and efficiency.
- Fail to achieve final integrate of different work groups.
- Timeouts for debugging and final deliverable.
- The risk of potential dissatisfaction from the users
- The risk of failing to respond to the feedback from the user survey



## Risk Management:

- Review work every 2 weeks. If any progress is falling behind, try: switch methods/ add workload/ adjust team structure...
- Conduct first user survey early (week 5), so we can have sufficient time to response to survey results.
- Communicate with client and tutors timely.

# Tools & Constraints

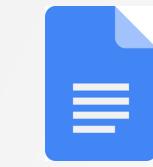
## Development Tools



Github



Python3



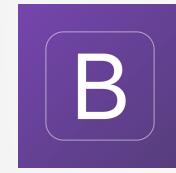
Google Doc



MongoDB



Flask



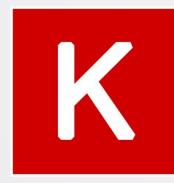
Bootstrap



Google Form



Gensim



Keras



Pytorch



Tensorflow

## Communication Tools



Slack  
With Client



WeChat  
With Team members

- In-Person Client Meeting every 2 weeks.
- Online Client Meeting all other weeks.
- Team Meeting whenever necessary.

## Process Management Tool



Trello

<https://trello.com/b/vxJ6hSCv/news-aggregator-sem2>

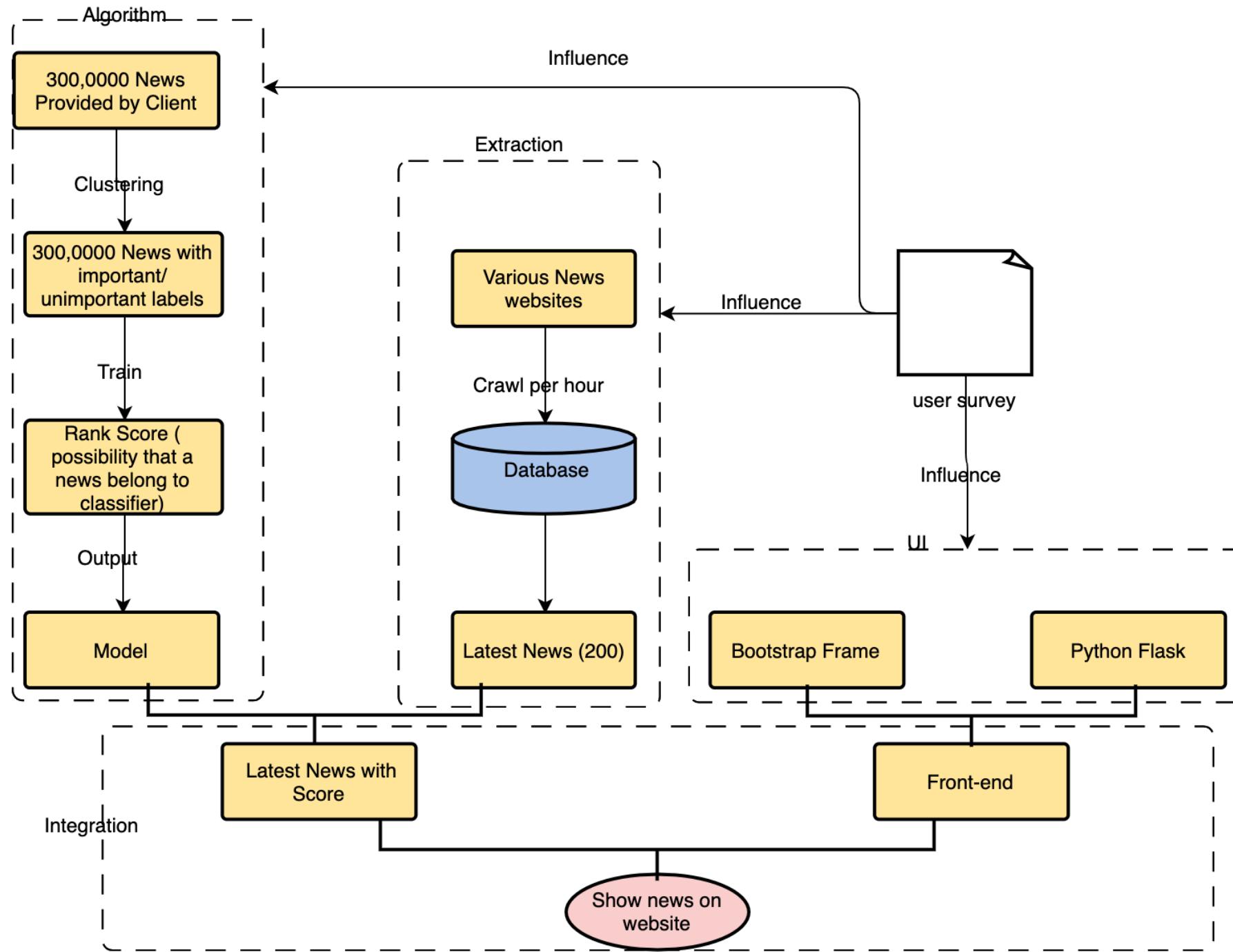
# Teamwork

Name	uid	Technical Role	Progress Management Role
Jiahua Liang	u6162679	UI & User	Spokesperson 1
Luokun Gong	u5917339		Minutes Taker
Xiangyun Kong	u6556183		Progress Tracker
Yulinag Zhang	u6782445	Extraction & Algorithm	Spokesperson 2
Vishnu Vardhan Jasti	u6611697		Clerical Assistant
Jun Yang	u6767560		Clerical Assistant
Chen Zhang	u6745297		Progress Tracker

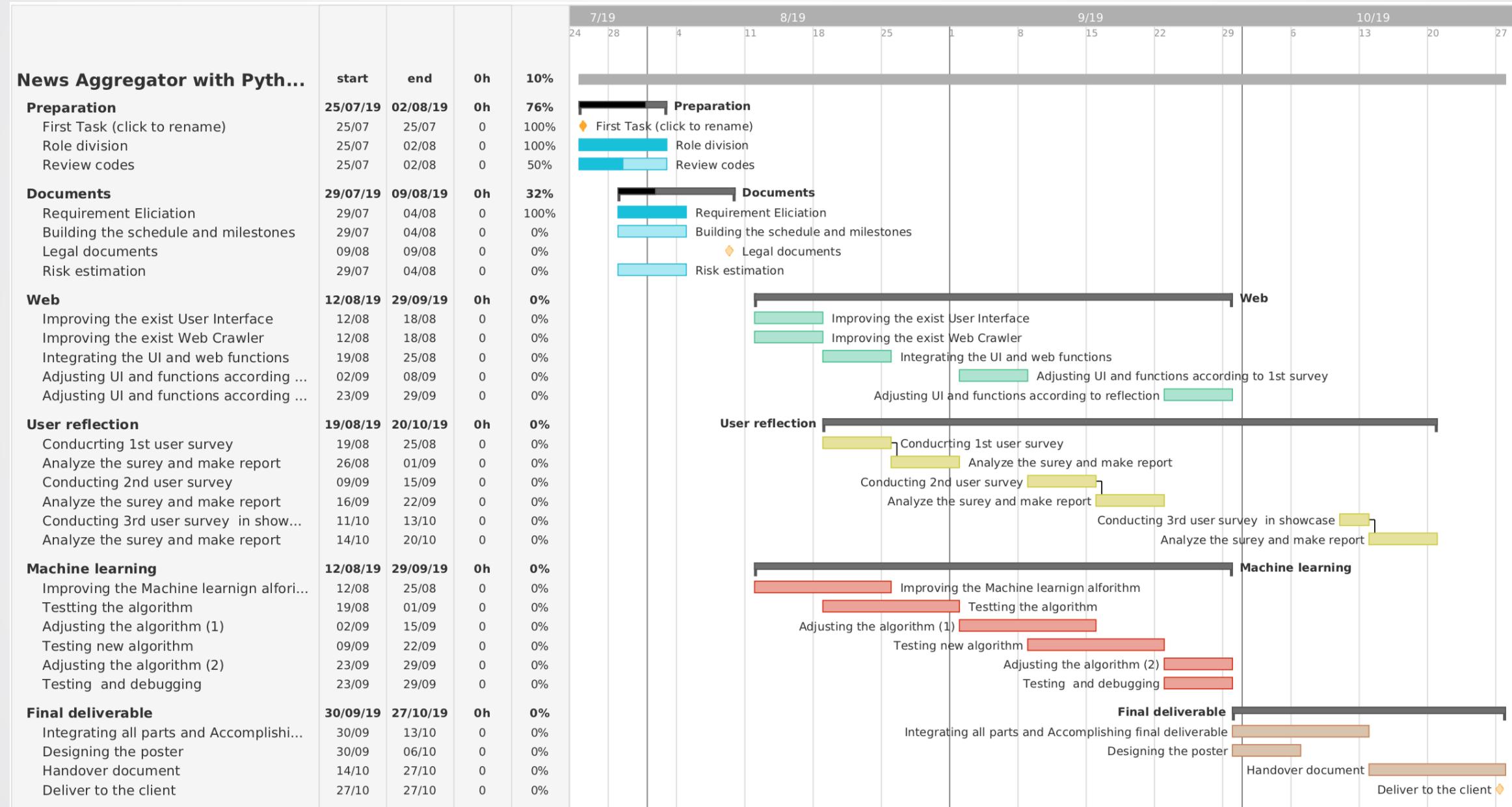
# Schedule

Week 1-2	<ul style="list-style-type: none"><li>• Form Team</li><li>• Divide Roles</li><li>• Set up development environment</li></ul>	Week 7-8	<ul style="list-style-type: none"><li>• Second user survey</li><li>• Reflect to user survey results</li></ul>
Week 3-4	<ul style="list-style-type: none"><li>• Improve extraction work</li><li>• Improve algorithm</li><li>• Improve UI</li></ul>	Week 9-10	<ul style="list-style-type: none"><li>• Poster</li><li>• Final deliverable</li><li>• Showcase (3rd user survey)</li></ul>
Week 5-6	<ul style="list-style-type: none"><li>• First user survey</li><li>• Reflect to user survey results</li></ul>	Week 11-12	<ul style="list-style-type: none"><li>• Handover documents</li><li>• Report to client</li></ul>

# Workflow



# Gantt chart



# User Story Point Matrix

Story points	User story
5	<ul style="list-style-type: none"><li>• As a visitor, I would like to select the news so that I can read its details on its original site.</li><li>• As a visitor, I would like to click the buttons of thumb up or thumb down so that I can decide to recommend it to others or not</li><li>• As a visitor, I would like to see the comments and rating of the news so that I can know others' feeling about the news.</li></ul>
10	<ul style="list-style-type: none"><li>• As a visitor, I would like to browse the Canberra news list, so that, I can know the events happening in Canberra that more likely to be related to me.</li><li>• As a visitor, I would like to browse the Australian news, so that I can know the important events happening in the Australia.</li><li>• As a visitor, I would like to browse the world news, so that I can know the important events happening in the world.</li><li>• As a visitor, I would like to write texts to comment on the news, so that I can describe my feeling about the news more specifically.</li></ul>
15	<ul style="list-style-type: none"><li>• As a visitor, I would like to search the specified news by keywords, so that I do not need to waste time to scroll down the news list to find the news.</li></ul>
30	<ul style="list-style-type: none"><li>• As a visitor, I would like to only read the most important news at the top of their lists, so that I can avoid wasting time on looking for the most important news.</li></ul>

# User Story Map

## NEWS AGGREGATOR

Recommending

Searching

Browse website

Review

Sorting news

Search the news

Browse the Global news and Canberra news

Select the news

Rating

Commenting

### MVP

Sorting news by importance

Search by key words

Browse the recommended Canberra news list

Select the news

Click the button of thumb up or thumb down

Commenting with texts

Browse the recommended global news list

See the rating and comments of the news

Browse the recommended Canberra news list

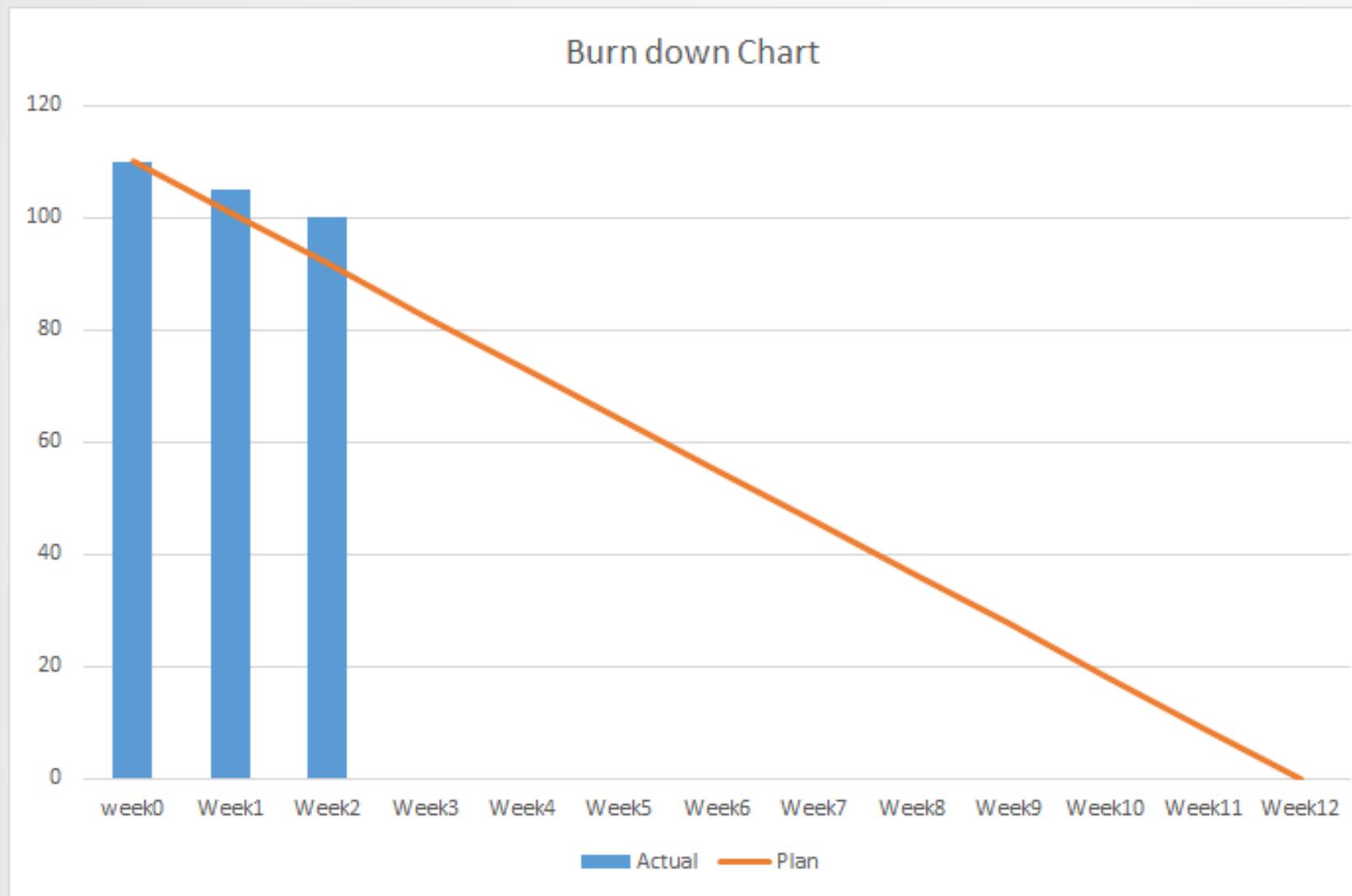
### Release 2

Sorting news by tags

Search by contents

Commenting with emotic icons

# Burndown Chart



The Burndown Chart is calculated according to our Trello record:

<https://trello.com/b/vxJ6hSCv/news-aggregator-sem2>



A minimalist, abstract background featuring light gray, semi-transparent geometric shapes like triangles and rectangles. A large, solid black rectangle is positioned above the text, and a smaller, solid black rectangle is centered below it. The overall aesthetic is clean and modern.

# Thank you