



Assessing partisan niche market strategies of media outlets using machine learning techniques

MSU Cloud Computing Fellowship 2022-2023

Presentation by Suhwoo Ahn, Doctoral student in Communication

Project description (1)

- Investigate what strategies political media outlets have adopted to attract an audience.
- Background
 - Media outlets show liberal or conservative positions to attract citizens' attention in creating political news (Hamilton, 2004; Mullainathan & Shleifer, 2005; Gentzkow & Shapiro, 2010).
 - Latecomer media outlets adopted a partisan niche market strategy – take extreme partisan positions to attract new customers.



Project description (2)

- Research questions
 - RQ1: Are latecomer media outlets more polarized in the news about presidents than established media outlets?
 - RQ2: Have established media outlets been polarized or depolarized in making news about presidents over time?
- Collected and analyzed 598,998 news articles from 32 media outlets in South Korea.
 - 17 established media outlets (2001 – 2018)
 - 15 latecomer outlets (2011 – 2018)



Employ cloud resources (1)

- Used Microsoft Azure and Google Colab.
- Using Microsoft Azure
 - Stored 1.6 million news articles (analyzed 0.6 million articles) in the storage in Microsoft Azure
 - Store and manage large size of data (texts)



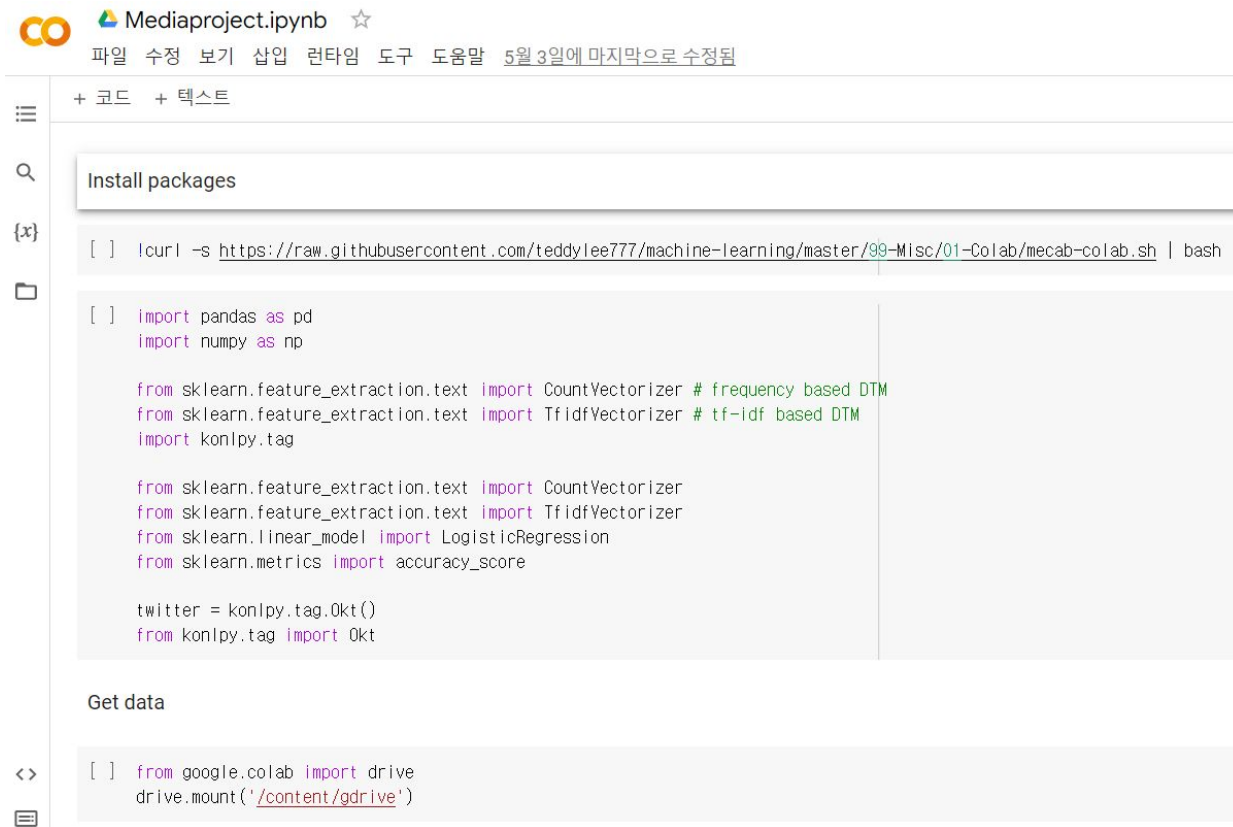
Employ cloud resources (2)

- Using Google Colab
 - Preprocess texts using natural language processing and create a document-term matrix.
 - Used a supervised machine learning technique (RASSO, RIDGE regression).
 - High computing power is needed to process natural language.



Employ cloud resources (3)

- Using Google Colab (example code)



The screenshot shows a Google Colab notebook interface. At the top, the notebook is titled "Mediaproject.ipynb" with a star icon. Below the title, there are links for "파일" (File), "수정" (Edit), "보기" (View), "삽입" (Insert), "런타임" (Runtime), "도구" (Tools), and "도움말" (Help), followed by the text "5월 3일에 마지막으로 수정됨" (Last modified on May 3). The notebook has two tabs: "+ 코드" (Code) and "+ 텍스트" (Text). The left sidebar contains icons for a menu, search, variables, and files. The main area displays the code in a light gray box. The code is organized into sections: "Install packages" and "Get data".

```
[ ] !curl -s https://raw.githubusercontent.com/teddylee777/machine-learning/master/99-Misc/01-Colab/mecab-colab.sh | bash
```

```
[ ] import pandas as pd
import numpy as np

from sklearn.feature_extraction.text import CountVectorizer # frequency based DTM
from sklearn.feature_extraction.text import TfidfVectorizer # tf-idf based DTM
import konlpy.tag

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

twitter = konlpy.tag.Okt()
from konlpy.tag import Okt
```

```
[ ] from google.colab import drive
drive.mount('/content/gdrive')
```

Success and challenges (1)

- Success
 - Store a large size of data and manage efficiently.
 - Overcome limits on my computer memory.
 - (Google Colab) Whenever I have an error, it show a relevant link to the StackOverflow website.



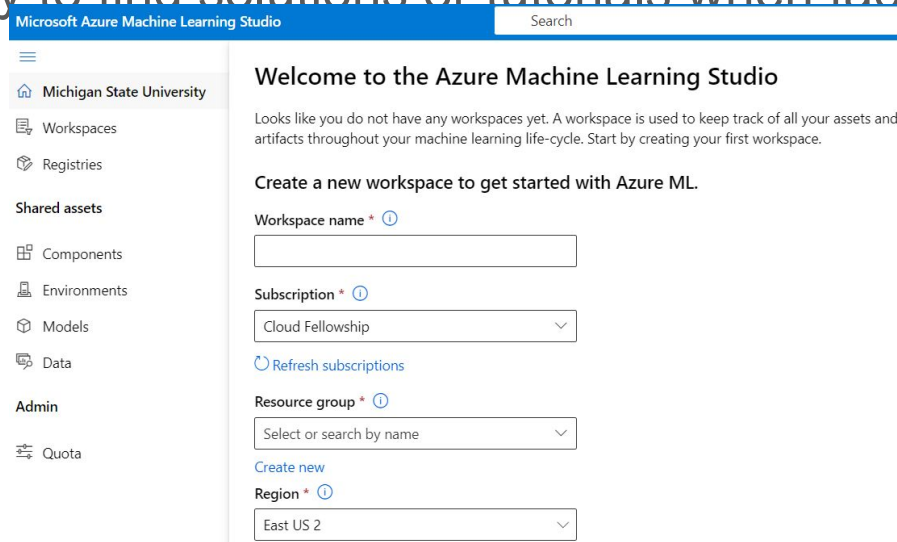
Success and challenges (2)

- Challenges
 - Have difficulties running Python codes in Azure – Followed a tutorial but could not solve the error.
 - Search for the Internet, but I could not find an answer.
 - Analyze Korean through *KoNPLy* on Python (Korean natural language processing tool)
 - *KoNPLy* is used in Python only and analyze nouns only.



Success and challenges (3)

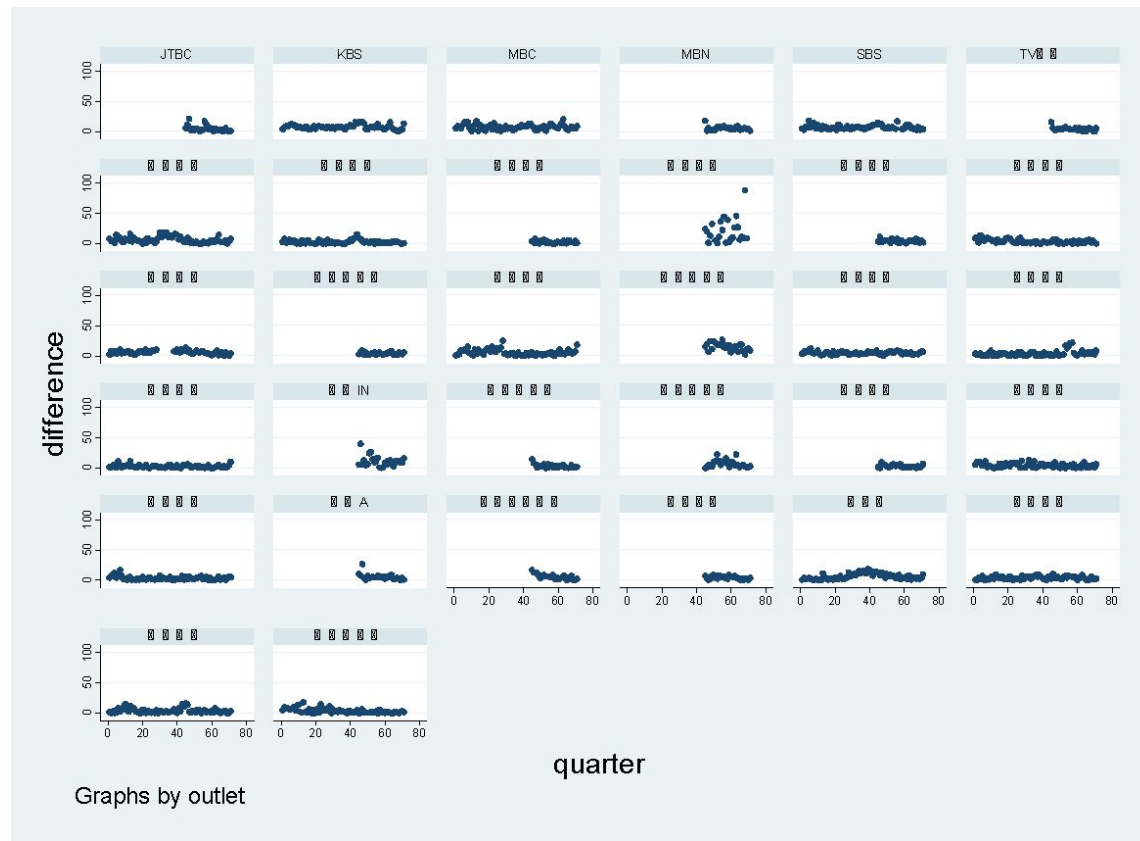
- Challenges
 - Azure is very professional but not easy to follow for those with basic levels of programming language.
 - Not easy to find solutions or tutorials when facing difficulties.



The screenshot displays the Microsoft Azure Machine Learning Studio interface. On the left is a navigation pane with a hamburger menu icon at the top, followed by 'Michigan State University' (with a house icon), 'Workspaces' (with a folder icon), 'Registries' (with a cube icon), 'Shared assets' (with a folder icon), 'Components' (with a grid icon), 'Environments' (with a server rack icon), 'Models' (with a cube icon), 'Data' (with a document icon), 'Admin' (with a gear icon), and 'Quota' (with a bar chart icon). The main content area has a blue header bar with 'Microsoft Azure Machine Learning Studio' and a search box. Below the header, it says 'Welcome to the Azure Machine Learning Studio' and provides a message: 'Looks like you do not have any workspaces yet. A workspace is used to keep track of all your assets and artifacts throughout your machine learning life-cycle. Start by creating your first workspace.' It then prompts the user to 'Create a new workspace to get started with Azure ML.' The form includes four required fields: 'Workspace name' (text input), 'Subscription' (dropdown menu with 'Cloud Fellowship' selected), 'Resource group' (dropdown menu with 'Select or search by name' selected), and 'Region' (dropdown menu with 'East US 2' selected). There are also links for 'Refresh subscriptions' and 'Create new'.

Outcome and future work (1)

- Changes in partisan tone of media outlets in South Korea.



Outcome and future work (2)

- Does the introduction of news cable channels (TV Chosun, JTBC, MBN, Channel A) in December 2011 affect polarization trajectory among media outlets?
- Conducted a group-specific linear growth model.



Outcome and future work (3)

- Fixed effects of GLGM (1)
 - Before introduction of new cable channels in December 2011, polarization levels among established media outlets remained stable ($b = .001$, $p = .962$).

difference	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
quarter0	.0006227	.0131353	0.05	0.962	-.0251221	.0263675
1.quarter_bi	3.06282	1.238496	2.47	0.013	.6354136	5.490227
quarter_bi#c.quarter0 1	-.0673433	.0234056	-2.88	0.004	-.1132173	-.0214692
_cons	5.264329	.5086403	10.35	0.000	4.267412	6.261245



Outcome and future work (4)

- Fixed effects of GLGM (2)
 - After introduction of new cable channels in December 2011, polarization levels among established media outlets decreased ($b = -.067$, $p = .004$).

difference	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
quarter0	.0006227	.0131353	0.05	0.962	-.0251221	.0263675
1.quarter_bi	3.06282	1.238496	2.47	0.013	.6354136	5.490227
quarter_bi#c.quarter0 1	-.0673433	.0234056	-2.88	0.004	-.1132173	-.0214692
_cons	5.264329	.5086403	10.35	0.000	4.267412	6.261245



Outcome and future work (5)

- Polarization of the established media outlets in South Korea remained stable before new cable channels were introduced in December 2011.
- After new channels were introduced, established media outlets were depolarized, which means that they show relatively neutral attitudes toward presidents.
- As latecomer media outlets show strong partisanship and receive more attention from strong partisans, established media outlets may strategically have neutral positions to appeal to less strong partisans.



Thank You

Contact Information:

Suhwoo Ahn

Doctoral Student

Department of Communication

Michigan State University

Email: ahnsuhwo@msu.edu

