

השוואת אסטרטגיות רגולטוריות בפיקוח בנקאי: ניתוח טיקסט במשמעות מדיניות



**הוגש במסגרת קורס כריאת טיקסט, בית הספר למנהל עסקים –
אוניברסיטת בר אילן**

מגיש: אריאל מתן

מרצה: ד"ר רוני רמון-גונן

הגדרת הבעיה ומטרות

רגולטורים בנקאים ברחבי העולם מפרטים מסמכי מדיניות אסטרטגיית שמאגדירים את תחומי הפעולה המרכזיים שלהם ואת סדר העדיפויות הפיקוח על הבנקים. מסמכים אלה הם מקור ידע חשוב אך בשל אורכם, ניסוחם הטכני והבדלי תחומי אחראיות קשה לנתחם באופן שיטתי וערוך השוואה מסודרת ביניהם.

בעידן של רגולציה שצריכה להיות דינמית ומערכות פיננסית משתנה בקצב מהיר, עולה הצורך של רגולטורים ומוסדות בנקאים ללמידה מהנעשתה במדינות אחרות. להבין מהם הסיכוןים שזוכים לשיקור והתייחסות רחבה, מהם הנושאים להם ניתנים משאבים גדולים יותר, וכי怎ן משתנה מיקוד הפיקוח בין רגולטורים שונים. ניתן כמו בעבודה זו יכול לסייע בגיבוש מדיניות מקומית רלוונטית תוך זיהוי מגמות בינלאומיות והפקת לקחים מניסיונים של אחרים.

מטרת עבודה זו היא "ישום טכניקות של כריית טקסט, על מסמכי מדיניות של רגולטורים במדינות מתניות בעולם". השיטה מאפשרת זיהות מיקוד נושאים מרכזיים בתוך מסמכי האסטרטגיה, להשוות בין רגולטורים, ולגלות הבדלים בין הגישות האסטרטגיות אותן הם מאמצים.

שאלות המחקר:

אילו נושאים חמורים ועולים במסמכי האסטרטגיה של רגולטורים שונים בעולם, וכי怎ן משתנה דגש האסטרטגיה בין מדינות?

האם ניתן לזהות מגמות משותפות או ייחודיות בהתמודדות עם סיכוןים ויעדים רגולטוריים?

סקרים בספרות

Exploring the Space of Topic Coherence Measures (Röder, Both & Hinneburg, 2015)

המאמר בודק מספר דרכי לבודוק האם נושא שנוצר על ידי Topic Modeling טוב להבנה ולקראיה אנושית. הוא מציע מדרדים שעובדים טוב יותר, ומראים איך ניתן לבחור את מספר הנושאים (K) בצורה טובה יותר. המאמר עוזר לבחור ולהסביר את ההחלטה על K מסוים ולהראות שהנושאים יצאו קוהרנטיים והגיוניים.

LDAvis: A method for visualizing and interpreting topics (Sievert & Shirley, 2014)

המאמר מדבר על כל שונם בכיתה ונעשה בו שימוש בעבודה זו, הכל מציג את הנושאים כבועות על מפה ומוסיף לכל נושא את המילים הכי חשובות בו. ישנה גם אופציה לאזן בין מילים שכיחות למילים ייחודיות. בצורה זו קל להבין על מה כל נושא מדבר ולתת לו את השם המתאים.

Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication (Hansen & McMahon, 2016)

המאמר בוחן את ההשפעה העצמאית של תוכן החלטת הריבית של בנק המركزي בארה"ב (Federal Reserve) בנפרד מההחלטה עצמה. החוקרים מנסים להפריד בין ההחלטה עצמה – הعلاאת, השארת, הורדת הריבית ביחס לצפויות השוק לבין "אזור טקסטואלי" שהוא המסר וניסוח ההודעה עצמה, שיכולה לרמז על החלטות עתידיות. החוקרים מראים באמצעות מודלים כמותיים怎么说 כאשר הריבית משתנה או אינה משתנה צפוי, המילים עצמן יכולות לגרום לתוצאות משמעותיות בשוקים ולהשפע על הכלכלת. המסקנה היא שלtekסטים רגולטוריים בפני עצם ישנה השפעה על השוק, لكن יש ערך בניתוח שלהם כגורם נפרד.

Transparency and Deliberation within the FOMC (Hansen, McMahon & Prat, 2018)

בדיקה של תמלili הוועדה הפדרלית לשוק פתוח בארה"ב מראה שככל שהשקיפות גדלה (למשל פרסום וידאו או תמלול של הדינום), החברים נוטים לדבר בצורה דומה יותר ופחות להביע דעתות שונות. שימושיים מסמכי אסטרטגיה בין רגולטורים/מדיניות, צריך לזכור שככל שקיימות יכולות לשנות את הסגנון והתוקן – זה יכול להסביר דמיון או שונות בנושאים. עם זאת, לאחר שאין לנו מידע על תמלולים או אינטראסים נוספים של המשתתפים, אין למצאי זה יישום ממשמעותי בעבודה הנוכחיית.

נתוניים

הנתוניים הרצויים

לצורך המענה על שאלת המחקר, האידיאל היה לאסוף מאגר של מסמכים אסטרטגיה רגולטורית עדכניים של רגולטורים מובילים בתחום הפיקוח על הבנקים ובעלי מבנה קרוב יחסית למבנה הרגולציה בישראל. המאגר הרצוי צריך לכלול מסמכים באנגלית בלבד אשר בהם תיאור מפורט של יעד הרגולטור, תחומי הסיכון שבהם הוא מתמקד, והצעדים המתוכננים לטווח הקרוב והבינוני. בנוסף, רצוי שהייתה מטא-דאטה על כל מסמך הכלול את שם הרגולטור, המדינה אליה הוא שייך, שנת פרסום לכל מסמך.

הנתוניים בפועל

הנתוניים שנאספו בפועל מרכיבים ממסמכים אסטרטגיה רגולטורית שהתקבלו מאתר אינטרנט רשמי של רגולטורים בנקאים שונים, ביניהם:

- בריטניה – (PRA) Prudential Regulation Authority
- האיחוד האירופי – (ECB) European Central Bank
- ארה"ב – (OCC) Office of the Comptroller of the Currency
- ארה"ב – (FDIC) Federal Deposit Insurance Corporation
- קנדה – (OSFI) Office of the Superintendent of Financial Institutions

פערים בין הנתוניים הרצויים לנאספים

- כלל הרגולטורים מתעסקים בפיקוח על הבנקים. עם זאת, עקב מבנה רגולטורי שונה בין מדינה לממשלה תחומי האחריות של הרגולטורים אינם זהים לחולוטין. ישנו רגולטור אחד "משלימים זה את זה" וישנו רגולטור בעל אחריות מלאה לפיקוח על הבנקים באופן לפחות לפיקוח על גופים פיננסיים אחרים.
- טווח השנים אינו אחיד – עבור חלק מהרגולטורים קיימים מסמכים עדכניים מאוד, בעודם שעבור אחרים זמינים רק מסמכים בני מספר שנים. החדש ביותר – 2025, הישן ביותר – 2022.
- חלק מהמסמכים כוללים טבלאות ותרשימים שהומרו לטקסט פשוט.
- ישנו הבדלים באורך ובפירוט בין המסמכים, מה שעלול להשפיע על משקלם בחישובים שונים לאורך הניתוח.

המסד עצמו:

- שורה = חלק מסמך אסטרטגיה – נושא, מטרה וכו'
- עמודות:
 - ID Global – מזהה ייחודי לכל מסמך.

- – שם המדינה. Country
- – שם הרגולטור. Regulator
- – שנת פרסום המסמך. Year
- – כותרת לטקסט. Title
- – גוף הטקסט הגלמי. Text
- – מזהה מסמך לפי רגולטור. Doc_ID

רלוונטיות לבעה

המסמכים שנאספו מהווים תיעוד רשמי של סדרי העדיפויות, ההערכות והאתגרים המרכזיים כפי שגדירים אותם רגולטורים בנקאים. ניתן טקסטואלי של מסמכים אלו באמצעות Topic Modeling מאפשר לזהות מגמות מסווגות ובדלים בין רגולטורים, ולספק תובנות מבוססות נתוניות לגבי התמకדות האסטרטגיית של הפיקוח על הבנקים בעולם.

ארגון וניקוי טקסט

הארגון והניקוי נעשה בסדר הבא:

1. **הסרת קישורים וכתובות דוא"ל** – אין מידע רלוונטי לתוכן אותו חיפשתי בעבודה.
2. **הסרת סימנים שאינם אותיות** – אין מסימעים לנתח ואפ מוסיפים רעש, הסרת רווחים תקל על בניית DTM טוב.
3. **הפקת עמודת טקסט לקובץ** – על להמשיך בתהליך הניקוי על ידי `tm_map` בצורה טובה.
4. **הקטנת אותיות** – למניעת כפילויות של מילים וביצוע ספירה מדויקת יותר
5. **Lemmatization** – לאחר התלבבות בין אופציה זו ל- stemming, העדפתני לנסוט לשומר על המשמעות ועל הדיווק. בנוסף, הבחירה בלטיניזציה תגרום להציג תוצאות מובנות ומוסדרות יותר.
6. **הסרת מילים עצירה** – הוסרו מילות ברירת המחדל בפונקציה "stopwords". הוסרו מילים נפוצות במסמכים רגילים וスペציית במסמכים רגולטוריים/אסטרטגיים כמו: "ensure", "consider" ועוד. בנוסף הוסרו שמותם של הרגולטורים שציינו הרבה את שם מסמך האסטרטגיה.
7. **הסרת סימני פיסוק** – סימני הפיסוק אינם מօסיפים הקשר ואפ יפריעו לייצרת DTM/TDM.
8. **הסרת ספרות** – על נמת לא ליצור הטיית תדיות או לשבש את תהליך ה- Topic Modeling. אומנם בחילק מהמסמכים הופיעו מספרים שמטרתם להמחיש מגמות אך רציתי להתמקד בתוכן לשוני.
9. **רווחים נוספים** – מיותרים וגורמים לרעש.

משתנים (פיצ'רים)

ייצוג טקסט – Bag of Words

הטקסט יוצג כ-BoW של Bigrams ו-Unigrams לאחר ניקוי ולמטיציה. לכל מסמך נספרו הופעתו מונחים ובננה DTM. למודל LDA נעשה שימוש בספירה גולמית של Unigrams.

סינון אוצר מילים

הוסרו Stopwords כלליות, מילים בהם יש שימוש גדול במסמכים רגולציה ואסטרטגיה, וכן מספרים ופיסוק. בוצעה למטייצה לאיחוד נטיות לשוניות. הסינון נדרש לצמצם רעש ולהציג מונחים נושאים המשקפים תכנים רגולטוריים.

פיצ'רים מבוססי נושא

לאחר התאמת מודל LDA עם 7 נושאים, לכל מסמך הופקה התפלגות נושאית: שבעה מספרים בין 0 ל-1 שסכומם 1 המשקפים את חלקו היחסי של כל נושא במסמך. בקטורי ההתפלגות הללו השתמשתי כפיצ'רים להציג: חישבותי ממוצעים לפי מדינה כדי ליצור פרופיל נושאים לכל אחת. והציגו אותן במפת חום להשוואה ויזואלית בין הקבוצות.

פיצ'רים מבוססי FinBERT

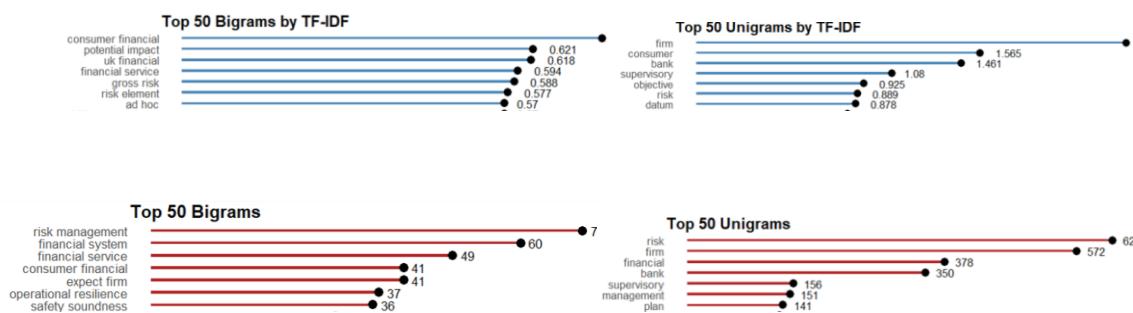
יצרתי Embeddings לכל מסמך באמצעות FinBERT. על בסיס מטריצת דמיון (קוויננס) בין המדינות חשבו: ממוצע דמיון של קרן מדינה לשאר המדינות, הזוג בעל הדמיון הגבוה ביותר, טווח גבהים בדנדוגרמה כמדד פיזור כללי, שירא לאשכולות על בסיס הדנדוגרמה. בנוסף, הציגו לאחר מכן דוגמאות של Unigrams + Bigrams אופייניים לכל בלוק לצורך הפרשנות שלו לתוצאות.

מטרה-משתנים לחיתוך

מדינה שימשה כמטרה חתך להציג התפלגות אך לא הוזנו למודל כפיצ'רים.

שיטות הניתוח ותוצאות

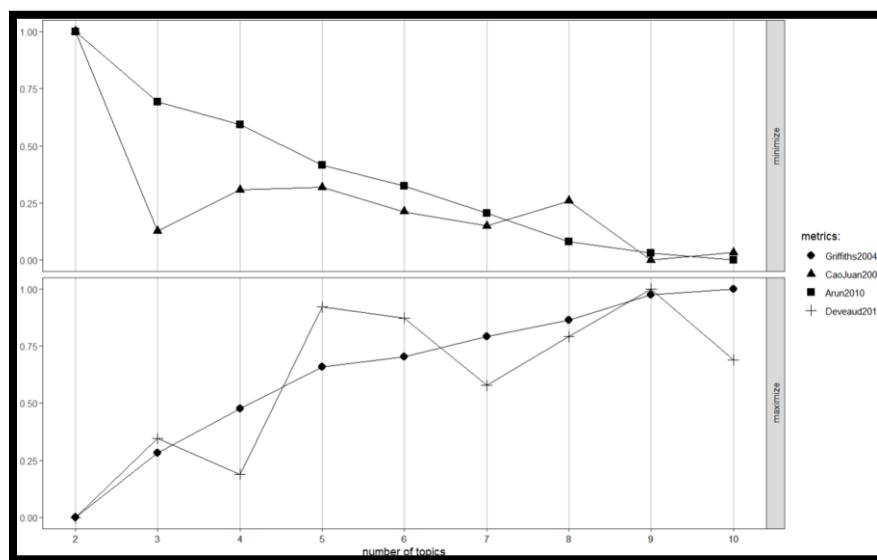
בנו מטריצות TDM + DTM ביוניגראם ובייגראם בשתי שיטות – BoW ו-TF-IDF והשוotta התפלגות המונחים בין השיטות. מצורפות טבלאות חלקיות – Top, הטבלאות המלאות מצורפות בסופחים.



לאחר שראיתי כי ישנם מילימ בעלות או השרוש שחוזרות על עצמן שוב ושוב הוחלט לבצע למטיציה ולא סטמינג על מנת להבין ולהציג בצורה טובה יותר את המילימ, מה שגרם לשינוי דרמטי במשקל המילימ, מילימ חדשים שנכנסו לטבלת ה- 50 Top ולשינוי משמעות בהשוואה במספר הנושאים ב- LDA. התוצאות המציגות למעלה הן לאחר למטיציה.

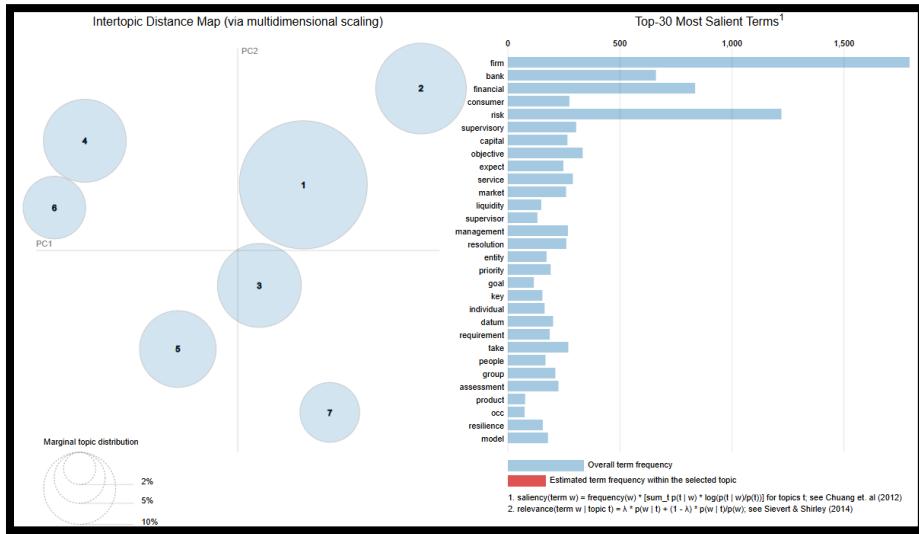
טבלאות 50 Top לפני ואחרי למטיציה נמצאות בסופחים.

השוואה מספר נושאים



נבחר $K=7$ לאחר הרצה על טווח של שנים עד עשרה נושאים. בשבע נושאים ישנה התקנסות טוביה בין המדדים למעט Deveaud2014. בחירה בתשעה נושאים יכולה לשפר עוד קצת את המדדים אך $K=7$ נבחר על מנת להוציא מודל פשוט וקל יותר לפרשנות.

מפת הנושאים (LDAvis) K=7



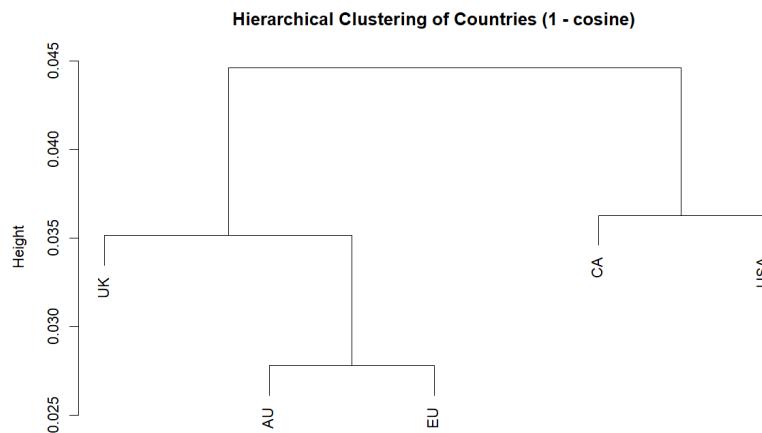
התרשימים מציג מפה הממחישה את מרחק הנושאים על מפה דו ממדית על ידי הפחיתה ממדים. הנושא בגודל ביותר הוא "ניהול סיכון ותכנון כשלים". על פי תרשימים נראה שיש בידול טוב בין כלל הנושאים למעט נושאים 4 ו-6, "פיקוח ואסטרטגיה" אל מול "רגולציה ומדיניות פרודנציאלית" שנראים קרובים אבל אינם חופפים. מה שמצודק שמות נפרדים. התמונה מראה נושאים קוגניטיביים ללא כפיפות, ומהזקפת את הבחירה בשבועה נושאים. הוסף נושאים יכולה ליצור מרכיבות והפחיתה נושאים תיזור תמונה פחותה ברורה על המתרחש בפועל.

מפת חום- השוואת בין מדינות לפי נושאי מיקוד



המפה מציגה את התפלגות הנושאים לפי מדינה, כפי שנגזרה ממודל LDA. השורות הן מדינות, עמודות הן נושאים. כל תא מציג את האחוז של אותו נושא מתוך כלל הטקסטים של אותה מדינה. גוון כהה יותר = חלק גדול יותר. בתאים נמצאים הערכות המדויקים באחוזים כאשר כל שורה מנורמלת ל-100%.

קלאסטרינג מדינות לפי דמיון קוסינוס באמצעות FinBERT



מודל השפה שנבחר הוא FinBERT, זו גרסה של מודל BERT שעבירה אימונים על מסמכים פיננסיים כמו דוחות כספיים וכתבות פיננסיות, لكن הוא מתאים יותר לעולם הבנקאות והרגולציה. הוא מתאים לניטוח בעבודה זו מכיוון שהוא יודע לפרש משמעות בהקשר פיננסי. בנוסף, הוא מודל שלא נלמד בכיתה.

הניתוח נעשה על ידי יצירת Embeddings מהקורפוס בשפת פיתון, ולאחר מכן ניתן שלהם בשפת R בסוף קובץ הקוד הראשי (קובץ הפיתון יוצר כנספח).

הדנדוגרמה מציגה קלאסטרינג של המדינות.

ציר Y: 1 – cosine, המרחק שבו אשכולות מתחברים. חיבור נמוך יותר = דמיון טקסטואלי גבוה יותר.

ציר X: תצוגתי, אין משמעות ממשית.

- ממוצע דמיון לשאר: CA 0.956 ,EU 0.955 ,AU 0.968 ,UK 0.961 ,USA 0.959
- הציג הדומה ביזנה: איחוד אירופי – אוסטרליה ($\cos = 0.972$)
- טווח דמיון כללי (Height): 0.94 – 0.97 , כולם דומים לכולם באופן יחסית.
- שני בלוקים עיקריים: USA-CA , UK-EU-AU ,

הסקת מסקנות

מפת החום:

- **הבנק האירופאי המركזי** – מתמקד באופן מובהק בפיקוח וסטרטגיה. ניתן להסביר את זה מנקודת תפקודו כבנק האחראי על כמות גדולה מאוד של בנקים והוא אף משמש כמפקח של מפקחים.
- **בריטניה** – מיקוד חזק בניהול סיכון ובשירותים פיננסיים. לאחר הפרישה של בריטניה מהאיחוד האירופי המדינה מנסה למשוך השקעות על ידי מיצב שוק אטרקטיבי, תחרותי, ובעיקר יציב.
- **ארה"ב** – מיקוד חזק בהגנת לקוחות, ציות ומערכות קהילתיות. בארה"ב ישנו מבנה רגולטורי מסווג יחסית אשר מחלק בו קיימים מספר גופים שונים. לאחר ונלקחו מסמכים של שני רגולטורים אחד מהם אחראי על תחום הגנת הצרכן (CFPB), יש להניח שנוצרה הטיה לכיוון נושאים אלה שאינם עומדים במרכז סדר היום של רוב הרגולטורים האחרים בתחום.
- **קנדָה** – מעורבות קהילתיות, רגולציה ו מדיניות פרודנציאלית (יציבות). היחידים שנဏנו דגש על גיוון, אקלים ועוד.
- **אוסטרליה** – רגולציה ויציבות.

קלאסטרינג עפ"י מدد קוסינוס באמצעות FinBERT:

פרשנות לgraf: הבסיס המשותף לכל הרגולטורים בסקטור הפיננסי הוא יציבות – ניהול סיכון ורגולציה פרודנציאלית. ההבדלים בין המדיניות בדנדוגרמה נובעים מהדגשים האחרים/משנים. האיחוד האירופי ואוסטרליה עוסקות כמעט אך ורק בנושא היציבות, ניהול סיכון, נזילות, תעוז, מבחני לחץ. ארה"ב וקנדה מוסיפות לאסטרטגיה נושאים אחרים כמו הגנת הצרכן, הוגנות, גיוון והכללה, מעורבות קהילתיות. בריטניה מדברת בעיקר על יציבות אך מדגישה גם תחרותיות וחדשנות – מה שמרחיק אותה מבחןת קרבה אל האיחוד האירופי ואוסטרליה.

דוגמאות תוכן – מה שמרחיק אותה מבחןת קרבה אל האיחוד האירופי ואוסטרליה:

EU/AU/UK – Prudential/Risk: “prudential”, “capital”, “liquidity”, “risk management”, “stress testing”.

USA/CA – Consumer/Community: “consumer protection”, “disclosure”, “fair lending”, “complaints”, “DEI/diversity”, “community engagement”.

מחקר המשך

- ניתוח ה-LDA בעבודה זו נעשה באמצעות Unigrams משיקולים של אמינות הספרות, ניסיון לייצור TDM/TDM פחות מדויל, ושימוש נושאים ברור וקל יותר. המשך המאמר יהיה שימוש ב-Bigrams.
- התוצאות עם גורם מקצועני על מנת לבצע נרמול טוב יותר של המנסכים. בישראל למשל, חטיבת הפיקוח על הבנקים בנק ישראל עוסקת בכלל הנושאים הקשורים לפיקוח על הבנקים, לעומת מדיניות אחרת בהן בעודה זו היא רק חלק מתפקידו של רגולטור אחד או שאותה העבודה מבוצעת על ידי מספר רגולטורים. הנרמול מהווה אתגר ממשמעותי.
- הרחבת מסד הנתונים כך שיכלול רגולטורים ומדיניות נוספת על מנת ליצור תמונה רחבה ואמיננה יותר, כולל את ישראל.
- בעבודה זו הוחלט לצמצם את מספר הנושאים כך שייהי מפורט אבל לא מבולגן. אך על מנת לראות תמונה מצב חדה יותר, לאחר הוספה מסמכים נוספים, נרחיב את מספר הנושאים.

אני חולב שהמחקר בשלב זה, הגיע ליעדים הנדרשים למעט ניתוח בין רגולטורי עקב כך שלא הי מוסיף רגולטורים לכל מדינה. כאשר אוסיף עוד רגולטורים אוכל לעשות ניתוח בין רגולטורי טוב, בשלב זה לא ניתן מענה למטרת מחקר זו. האתגרים המרכזיים היו בبنית מסד הנתונים בצורה טובה, נרמול הנתונים בתוך המנסכים עצם וניסוי לכטוט את כל תפקידי הפיקוח על הבנקים ושימוש הנושאים שלקח קצת זמן ולא היה ברור ממבט ראשון.

תרומת המאמר:

תרומת המאמר היא יצירת מפה השוואתית של המיקוד הרגולטורי בין רגולטורים מוביילים בעולם, שמבליטה במה כל רגולטור מתמקד ומה חסר אצל אחרים—וכך מאפשרת לאתר תחומיים שכדי לאמץ או לחקז. בנוסף, המאמר מספק צפיה קידמה בזיהוי מגמות מתחווות והתפתחויות צפויות ברגולציה הפיננסית.

מקורות

חומר שנלמד בכיתה: קוד, מציגות, החלטות.
נעזרתי ב- ChatGPT, בעיקר בשימוש הנושאים, תיקון קוד, וניתוח באמצעות FinBERT זה בשפת R והן בשפת פיתון לייצרת ה- Embeddings.

מסמכים האסטרטגיה של הרגולטורים עצם:

RPA: https://www.bankofengland.co.uk/prudential-regulation/publication/2024/april/pra-business-plan-2024-25?utm_source=chatgpt.com

ECB: https://www.banksupervision.europa.eu/framework/priorities/html/ssm.suppervisory_priorities202212~3a1e609cf8.en.html?utm_source=chatgpt.com

OCC: https://www.occ.gov/publications-and-resources/publications/banker-education/files/pub-occ-strategic-plan-2023-2027.pdf?utm_source=chatgpt.com

FDIC: https://www.fdic.gov/analysis/risk-review/2024-risk-review/2024-risk-review-full.pdf?utm_source=chatgpt.com

OSFI: https://www.osfi-bsif.gc.ca/en/about-osfi/reports-publications/osfis-annual-risk-outlook-fiscal-year-2024-2025?utm_source=chatgpt.com

ליקם של המאמרים בסקרת הספרות עזרו לחזק החלטות מסוימות:

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)* (pp. 399–408). ACM.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Association for Computational Linguistics.

Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99(S1), S114–S133

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC. *The Quarterly Journal of Economics*, 133(2), 801–870.