## تحلیل احساس متون خبری و بررسی نقش آن‌ها در وضعیت نقدینگی بانک‌ها

در تحقیقاتی زیادی ثابت شده است که بازارهای مالی از اخبار رویدادهای سیاسی، اقتصادی و حتی ورزشی تاثیر می‌پذیرد. به گونه‌ای که خبر و میزان نفوذ آن باعث ایجاد هیجانات و انتظاراتی مثبت و منفی شده که واکنش‌هایی را از سمت کنش‌گران مالی در پی دارد.

از آنجا که بانک یکی از اهرم‌های بزرگ در اقتصاد بوده و خود نیز کنشگر اقتصادی دسته‌بندی می‌شود، می‌توان انتظار داشت که اخبار مکتوب و روایی می‌توانند بر بانک اثر گذارده و مولفه‌ای همچون نقدینگی ماهانه‌ی بانک از مولفه‌هایی است که این آثار را بازتاب می‌دهد.

در این پژوهش علمی و عملی سعی بر آن است نوع و میزان تاثیر مشخص و با ابزارهای آماری بررسی شود. همچنین هدف دوم و البته مهم‌تر، یافتن خبرگزاری‌های موثر، نوع متون موثر و درنهایت کلید واژه‌های موثر در تغییرات نقدینگی می‌باشد.

## نقشه‌ی راه

قبل از شرح مسیر و روند کاری، شرحی مختصر بر تعریف خبرهای مورد استفاده الزامی بنظر می‌رسد. در این پژوهش اخبار مکتوب و نوشتاری استفاده می‌شود. بدین منظور که اخبار به کار رفته از نوع متن دیجیتال بوده و به زبان فارسی می‌باشند. لذا داده‌ها از خبرگزاری‌های آنلاین یا خبرگزاری‌های چاپی که نسخه‌ی دیجیتال نیز دارند جمع‌آوری می‌شود. موضوعات کلی اخبار در زمینه‌ی اقتصادی و سیاسی می‌باشد. زیرا که نگارندگان، در ابتدای راه، احتمال موثر بودن این دو نوع از اخبار را بیش‌تر می‌دانند. روند کلی کار به شکل مختصر در ادامه بیان می‌شود.

## جمع آوری متون خبری:

داده‌ی خام اولیه متون خبری هستند که با موضوعات مختلف که از خبرگزاری‌های برخط جمع‌آوری می‌شود. خبرگزاری‌های فارس نیوز، دنیای اقتصاد، تجارت نیوز و انتخاب گزینه‌های مناسبی برای این امر بنظر می‌رسند. از این بین و در مرحله‌ی اول از خبرگزاری فارس نیوز اخبار در موضوعات مختلف جمع‌اوری شده و ذخیره شده‌است. اخبار با موضوع برای شروع مورد استفاده قرار گرفت.

## **پردازش** اولیه و پاک‌سازی متون

متن یک نوع داده‌ی خام و غیرساخت یافته است. که علاوه بر خود متن ممکن است در آن از لینک، ایمیل، ارجاعات و اشکال استفاده شده‌باشد. متن ممکن است با فرمت‌های و ابزارهای مختلف نوشته شده باشد که به زبان اصلی متن مرتبط نباشد. این مشکل بصورت خاص برای زبان فارسی بیش‌تر است. به‌طور مثال هنگام نوشتن خبر ممکن است از اعداد انگلیسی بجای اعداد فارسی استفاده شود یا از اعداد و حروف عربی در نوشتار استفاده شود. موارد مذکور با استفاده از عبارت منظم و لغت‌نامه‌های جایگزینی، تشخیص داده‌شده و جایگزین یا حذف می‌شوند. در آخر این مرحله متن ساختار خود را داشته و مفهوم و متن تغییری نمی‌کند.

همچنین متن ممکن است شامل فاصله‌ی اضافی باشد یا نیم‌فاصله رعایت نشده باشد که در مورد اول با استفاده از عبارت منظم تشخیص و حذف می‌شود و در مورد دوم از توابع کتاب‌خانه‌ای همچون HAZM استفاده می‌شود. سپس با استفاده از semmer و lemetizer از کتابخانه‌ی HAZM استفاده می‌شود تا شکل‌های مختلف کلمات حذف شده و ریشه‌ی کلمات در متن استفاده شود. در این حالت تعداد کلمات متن کم‌تر شده و از نظر ریاضی تنوع کم‌تر می‌شود. به طور مثال کلمه‌ی « می‌رود» با کلمه‌ی «رفته‌است» به شکل یکسان و با ریشه‌ی این دو کلمه در متن استفاده می‌شود.

## دسته‌بندی کلی موضوعی متون

موضوعات سیاسی و اقتصادی بسیار عام هستند. ممکن است شکستن و گروه‌بندی این موضوعات کلی به بخش‌های کوچک‌تر تحلیل‌ها را دقیق‌تر و بهتر نماید.

## استخراج کلیدواژه‌های متن

متن حال حاضر پر از کلمات و اجزای مختلف در ساختار هر پاراگراف خبر و هر جمله است تا مفهوم را با استفاده از قواعد فارسی به خواننده صریح و واضح انتقال دهد. اما ماشین می‌تواند با توجه به کلید واژه‌هایی از متن و بدون خواندن کل متن مفهوم متن را دریافت کند. کاری که خواننده‌های ثابت و کارشناس نیز انجام می‌دهند.

برای استخراج کلیدواژه‌ها از متن روش‌هایی آماری همچون TF-IDF و N-gram می‌توان به کار برد.

همچنین می‌توان ازمدل‌های یادگیری ماشین از پیش آموزش داده‌شده همچون Bert، Parsbert، hazm و ... استفاده کرد و قسمت‌های بی‌معنی و کم ارزش جملات را حذف و سپس از روش‌های آماری برای یافتن کلیدواژه‌ها استفاده نمود.

سپس در مرحله‌ی بعدی متون توسط روش‌های جاساز از حالت متنی به عدد تغییر شکل داده و به ماتریسی از اعداد تغییر می‌یابد تا توسط مدل‌های یادگیری قابل فهم و پردازش باشد.

## اعتبارسنجی کلیدواژه‌ها و حذف کلمات پرتکرار

بعد از انجام مراحل بالا با توجه به جدید بودن کار امکان مقایسه‌ی کلید واژ‌ها با دیگر روش‌ها نیست. لذا با خوشه‌بندی کلمات و بررسی معیارهای خوشه‌بندی و همچنین با کاهش بعد به دو بعد قابل مشاهده توسط انسان و بررسی چشمی می‌توان میزان کیفیت کار استخراج ویژگی و روش تبدیل کلمات به اعداد را بررسی کرده و درصورت نیاز به تقویت این دو بخش پرداخت.

## تحلیل احساسی مبتنی بر کلیدواژه:

سپس داده‌ی موجود با استفاده از روش‌های تحلیل احساسی، قطبیت آن و میزان مثبت، خنثی و منفی بودن خبر محاسبه و امتیاز گذاری می‌شود.

در پایان این مرحله اخباری داریم که بصورت کلی حس آن‌ها مشخص شده‌است و از طرفی کلیدواژه‌هایی وجود دارد که از این متون استخراج شده‌است. ممکن است نگاشت این کلیدواژه‌ها یا کلمات چندگانه با حس متن بتواند شهودی مناسبی ارائه دهد که بتوان روابطی بین نقدینگی و این اخبار یافت.

## اعتبارسنجی براساس میزان نقدینگی بانکی

در این مرحله روندی ارائه می‌شود که بتوان روابط ادعایی در مرحله‌ی قبل را اعتبارسنجی کرد. بدین منظور از روش‌های آماری می‌توان استفاده کرد و با محسابه میزان هم‌بستگی یا هم اطلاعی و محاسبه‌ی p-value میزان اثرگذاری کلیدواژه‌ها و معتبر بودن اثر بررسی شود.

در روندی دیگر و در صورت استفاده از روش‌های سری زمانی می‌توان با استفاده از روال LOOCV ماه به ماه جلو رفت و هر ماه با استفاده از اطلاعات گذشته ماه بعدی را پیش‌بینی کرد. در این صورت دقت پیش‌بینی و نزولی بودن خطا می‌تواند تحلیل‌ها و روش‌ به‌کار رفته را از نظر اعتبار بررسی نماید.

## تحلیل و بررسی ارتباط بین نتایج الگوریتم و روند نقدینگی

در این مرحله و در پایان می‌توان ابزار و نرم‌افزاری ارائه داد که با دریافت اخبار به تحلیل آن‌ها پرداخته و براساس روند نقدینگی گذشته و اعمال ‌مدل‌های یادگیری و آماری به تجزیه و تحلیل پرداخته و روند یا مقدار نقدینگی در ماه آتی را پیش‌بینی می‌کند.