

# Computational Analysis of Persian Texts with a Focus on Religious and Mystical Literature: Introduction to Reliable Software and Virtual Platforms for Digital Text Analysis

Yahya Ataei<sup>✉</sup> 

1. Assistant Professor of Persian Language and Literature, Payame Noor University:  
Y.ataei@pnu.ac.ir

## Article Info

**Article type:**  
Research Article

**Article history:**  
**Received:** 13 May 2025  
**Received in revised form:**  
19 June 2025  
**Accepted:** 26 June 2025  
**Available online:**  
23 August 2025

**Keywords:**  
digital literature,  
distant reading,  
topic modeling,  
statistical stylometry,  
computational linguistics,  
text mining.

## ABSTRACT

With the remarkable growth of digital humanities in recent decades, new horizons have opened for the analysis of classical texts, particularly within Persian religious and mystical literature. This article aims to introduce contemporary theoretical approaches and analytical tools, highlighting the potential of digital literature to support Islamic and mystical studies. Unlike studies focusing on a specific corpus, this research seeks to present a range of specialized methods and software employed in leading global universities such as Stanford, Harvard, and MIT within the field of digital literary studies, explaining their capabilities for analyzing Persian texts.

The theoretical framework of the article is grounded in concepts such as distant reading (Franco Moretti), corpus linguistics, topic modeling using the LDA algorithm, and statistical stylometry. Accordingly, software such as Voyant Tools, AntConc, MALLET, Gephi, and Word2Vec are introduced, with an examination of their application in analyzing explicit and implicit concepts, stylistic structures, and semantic patterns in Persian texts. The findings indicate that the development of technical infrastructure—including corpus-building platforms and text-mining tools—can create a systematic transformation in literary studies. Computational approaches, by minimizing individual biases, enable large-scale analysis, comparative studies of styles, and integration of statistical methods with rhetorical readings.

Among the advantages of these methods are the discovery of hidden semantic patterns without imposing the researcher's preconceptions, visualization of concepts, and the ability to analyze long and diverse texts. These tools also have extensive applications in interdisciplinary research, particularly at the intersection of literature, linguistics, and computer science. However, challenges remain, such as the need for clean, digitized corpora, difficulties in interpreting statistical outputs, the lack of semantic lexicons for classical Persian, and technical complexities in preprocessing Persian texts. Therefore, the article emphasizes the necessity of collaboration among specialists in literature, computational linguistics, and Islamic studies to advance this type of interdisciplinary research.

**Cite this article:** Ataei, Y. (2025). Computational Analysis of Persian Texts with a Focus on Religious and Mystical Literature: Introduction to Reliable Software and Virtual Platforms for Digital Text Analysis. *Digital Islamic Studies and Humanities*, 1 (1), 355-380.  
<https://doi.org/10.22034/disah.2025.2076077.1055>



© The Author(s). **Publisher:** Research Center for Digital Islamic Studies and Humanities (RCDISAH).

**DOI:** <https://doi.org/10.22034/disah.2025.2076077.1055>

## تحلیل رایانشی متون فارسی با تمرکز بر آثار ادبی دینی و عرفانی

(معرفی نرم‌افزارها و بسترهای مجازی معتبر در تحلیل دیجیتال متون)

یحیی عطائی<sup>۱</sup> ✉

۱. استادیار زبان و ادبیات فارسی دانشگاه پیام‌نور، رایانامه: Yataei@pnu.ac.ir

## اطلاعات مقاله

## چکیده

نوع مقاله:

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۲/۲۳

تاریخ بازنگری: ۱۴۰۴/۰۳/۲۹

تاریخ پذیرش: ۱۴۰۴/۰۴/۰۵

تاریخ انتشار: ۱۴۰۴/۰۶/۰۱

کلیدواژه‌ها:

ادبیات دیجیتال،

خوانش دور،

مدل‌سازی موضوعی،

سبک‌شناسی آماری،

زبان‌شناسی رایانشی،

داده‌کاوی متون.

با رشد چشم‌گیر علوم انسانی دیجیتال در دهه‌های اخیر، افق‌های نوینی برای تحلیل متون کهن، به‌ویژه در حوزه ادبیات فارسی دینی و عرفانی گشوده شده است. این مقاله با هدف معرفی رویکردهای نظری و ابزارهای تحلیلی نوین، به تبیین ظرفیت‌های بالقوه ادبیات دیجیتال در خدمت پژوهش‌های اسلامی و عرفانی می‌پردازد. برخلاف پژوهش‌هایی که بر تحلیل یک پیکره خاص تمرکز دارند، این مطالعه درصدد است مجموعه‌ای از روش‌ها و نرم‌افزارهای تخصصی را معرفی کند که در دانشگاه‌های معتبر جهانی همچون استنفورد، هاروارد و ام‌آی‌تی در حوزه «مطالعات ادبی دیجیتال» به کار گرفته می‌شوند و قابلیت‌های آن‌ها را در تحلیل متون فارسی شرح دهد. چارچوب نظری مقاله بر مفاهیمی چون خوانش دور (Franco Moretti)، زبان‌شناسی پیکره‌ای (Corpus Linguistics)، مدل‌سازی موضوعی (Topic Modeling) با الگوریتم LDA و سبک‌شناسی آماری (Stylometry) استوار است. در این راستا، نرم‌افزارهایی نظیر Mallet، AntConc، Voyant Tools، MALLET، Word2Vec و Gephi معرفی شده‌اند و نحوه کاربرد آن‌ها در تحلیل مفاهیم پنهان و آشکار، ساختارهای سبکی و الگوهای معنایی متون فارسی بررسی شده است. نتیجه کلی پژوهش نشان می‌دهد که توسعه زیرساخت‌های فنی، از جمله پلتفرم‌های پیکره‌سازی و ابزارهای متن‌کاوی، می‌تواند تحولی روشمند در مطالعات ادبی رقم بزند. رویکردهای رایانشی با پرهیز از سوگیری‌های فردی، امکان تحلیل کلان، مطالعه تطبیقی سبک‌ها و تلفیق روش‌های آماری با خوانش‌های بلاغی را فراهم می‌سازند. از جمله مزایای این روش‌ها می‌توان به کشف الگوهای پنهان معنایی بدون دخالت پیش‌فرض‌های پژوهشگر، قابلیت تصویری‌سازی مفاهیم و توان تحلیل متون بلند و متنوع اشاره کرد. این ابزارها همچنین در پژوهش‌های میان‌رشته‌ای، به‌ویژه در تقاطع ادبیات، زبان‌شناسی و علوم رایانه، کاربرد گسترده‌ای دارند. با این حال، چالش‌هایی همچون نیاز به پیکره‌های پاک و دیجیتال‌شده، دشواری در تفسیر خروجی‌های آماری، کمبود واژه‌نامه‌های معنایی برای فارسی کلاسیک و پیچیدگی فنی در پیش‌پردازش متون فارسی همچنان پابرجاست. از این‌رو، مقاله بر ضرورت همکاری میان متخصصان ادبیات، زبان‌شناسی رایانشی و علوم اسلامی برای توسعه این نوع پژوهش‌های میان‌رشته‌ای تأکید می‌ورزد.

استناد: عطائی، یحیی (۱۴۰۴). تحلیل رایانشی متون فارسی با تمرکز بر آثار ادبی دینی و عرفانی. علوم انسانی و اسلامی

دیجیتال، ۱ (۱)، ۳۵۵-۳۸۰. <https://doi.org/10.22034/disah.2025.2076077.1055>



ناشر: پژوهشکده علوم اسلامی و انسانی دیجیتال (مرکز تحقیقات کامپیوتری علوم اسلامی نور). © نویسندگان.

## مقدمه

در عصر تحولات پرشتاب دیجیتال، علوم انسانی نیز دگرگونی‌هایی بنیادین را تجربه کرده است. این تحولات، از یک سو امکان بهره‌گیری از فناوری‌های رایانشی برای تحلیل متون و داده‌های فرهنگی را فراهم ساخته‌اند و از سوی دیگر، افق‌های نوینی برای روش‌شناسی، نظریه‌پردازی و بازخوانی منابع کلاسیک گشوده‌اند. یکی از مهم‌ترین دستاوردهای این تحول، ظهور شاخهٔ میان‌رشته‌ای «علوم انسانی دیجیتال» (Digital Humanities) است؛ حوزه‌ای که با تلفیق روش‌های کتی، زبان‌شناسی رایانشی، داده‌کاوی و تحلیل آماری، امکان بررسی ساختاری، معنایی و تاریخی متون انسانی را در مقیاس کلان فراهم می‌سازد.

ادبیات فارسی، به‌ویژه در شاخه‌های دینی و عرفانی، همواره حامل بار معنایی، الهیاتی و زیبایی‌شناختی غنی بوده است. با این حال، روش‌های سنتی تحلیل ادبی - که عمدتاً مبتنی بر خوانش نزدیک و تأمل فردی‌اند - در استخراج الگوهای پنهان معنایی، سیر تحول سبک‌ها و ساختارهای گفتمانی گسترده، با محدودیت‌هایی مواجه‌اند. این محدودیت‌ها در بستر نظام‌های آموزش از راه دور دوچندان می‌شود؛ چراکه در این نظام‌ها، درک مفاهیم عمیق دینی و عرفانی صرفاً از طریق متون دیجیتال و خواندنی صورت می‌گیرد و اغلب با نوعی فاصله‌گذاری معرفتی همراه است.

پژوهش حاضر در پاسخ به این نیاز، در پی ارائهٔ چارچوبی نظری و روشی برای تحلیل ماشینی متون دینی فارسی است؛ چارچوبی که با بهره‌گیری از نظریهٔ «خوانش دور» (Distant Reading) فرانکو مورتی و روش‌های مکملی چون مدل‌سازی موضوعی (Topic Modeling)، سبک‌شناسی آماری (Stylometry)، زبان‌شناسی پیکره‌ای (Corpus Linguistics) و مدل‌های واژه‌برداری معنایی (Word Embeddings)، امکان تحلیل ساخت‌یافته و معناشناختی متون عرفانی و دینی را فراهم می‌سازد.

پیشینهٔ این رویکرد عمدتاً در جهان غرب و در حوزهٔ ادبیات انگلیسی شکل گرفته است؛ در آثار پژوهشگرانی چون فرانکو مورتی، تد آندروود و متیو جوکرز، و در قالب پروژه‌هایی همچون Stanford Literary Lab و Mining the Dispatch در این پروژه‌ها، روش‌هایی نظیر الگوریتم LDA، مدل Word2Vec، تحلیل خوشه‌ای مفاهیم و ابزارهایی مانند Voyant Tools، MALLET، Gephi و AntConc برای کشف الگوهای زبانی، معنایی و سبکی در متون به‌کار گرفته شده‌اند.

با این حال، استفاده نظام‌مند از این روش‌ها در تحلیل متون فارسی اسلامی هنوز در مراحل آغازین قرار دارد. یکی از معدود پژوهش‌های ایرانی در این زمینه، مقاله‌ای با عنوان «بررسی و شناخت کارکردهای هوش مصنوعی در پردازش و تحلیل متون نثر فارسی بر پایه نثرپژوهی رایانشی» نوشته فیروزی مقدم (۱۴۰۳) است. این مقاله با تمرکز بر نثر فارسی، نقش ابزارهای هوش مصنوعی را در استخراج الگوها و مضامین زبانی و معنایی بررسی کرده و نشان می‌دهد که تحلیل رایانشی چگونه می‌تواند به مطالعات نسخه‌شناسی، سبک‌شناسی و پژوهش‌های تطبیقی در ادبیات فارسی یاری رساند.

پژوهش حاضر با تمرکز بر نیازهای آموزشی در نظام‌های آموزش غیرحضوری - و با تکیه بر تجربه نویسنده در دانشگاه پیام نور - در پی معرفی و امکان‌سنجی ظرفیت‌ها، ابزارها و روش‌های نوین تحلیل دیجیتال است. هدف نهایی، ارائه مدلی میان‌رشته‌ای برای تحلیل خودکار متون دینی فارسی است؛ مدلی که ضمن حفظ زیبایی‌شناسی متن، بتواند داده‌محور و ساخت‌یافته به فهم دقیق‌تر مفاهیم اسلامی در زمینه آموزش و پژوهش دانشگاهی کمک کند.

استفاده از ابزارهای نرم‌افزاری و سخت‌افزاری در تحلیل متون ادبی، موافقان و مخالفانی دارد و گاه با انتقاداتی نیز مواجه شده است. با این حال، هدف این مقاله صرفاً معرفی این شیوه‌ها و تبیین امکانات آن‌هاست، نه اتخاذ موضعی خاص در قبال آن‌ها. از سوی دیگر، طرح این مباحث می‌تواند در فضای علمی و فرهنگی ایران زمینه‌ساز گشودن باب گفت‌وگو میان طرفداران و منتقدان گردد؛ امری که به خودی خود می‌تواند گامی مثبت در جهت پویایی حوزه نقد و مطالعات ادبی به‌شمار آید. اهداف مقاله عبارت‌اند از: ۱. معرفی چارچوب‌های نوین در علوم انسانی دیجیتال برای تحلیل متون دینی فارسی؛ ۲. بررسی کاربرد ابزارهایی مانند LDA، Stylometry، Word2Vec و تحلیل شبکه مفهومی؛ ۳. تحلیل انتقادی قابلیت‌های این روش‌ها در آموزش از راه دور و تولید محتوای آموزشی ساختاریافته.

فرضیه اصلی پژوهش: روش‌های تحلیل معنایی رایانه‌محور می‌توانند در فهم عمیق‌تر مفاهیم اسلامی، کشف الگوهای سبکی و معنایی پنهان، و کاهش ابهام متون آموزشی فارسی - به‌ویژه در محیط‌های فاقد تعامل حضوری یا ساختار کلاسیک آموزشی - نقش مؤثری ایفا کنند.

این پژوهش از نوع نظری-کاربردی و با رویکرد کیفی-توصیفی انجام شده است و با تکیه بر تحلیل مفهومی و مطالعه اسنادی، به معرفی چارچوب‌ها، ابزارها و روش‌های نوین تحلیل دیجیتال در زمینه متون دینی فارسی می‌پردازد. تمرکز مقاله نه بر تحلیل محتوای یک متن خاص، بلکه بر تبیین ظرفیت‌های روش‌شناختی علوم انسانی دیجیتال و ارزیابی قابلیت‌های آن در تحلیل مفهومی و ساختاری ادبیات اسلامی در بستر دیجیتال است.

داده‌ها از طریق مطالعه کتابخانه‌ای، بررسی متون نظری، پژوهش‌های بین‌المللی و مستندات پروژه‌های علمی در حوزه‌های علوم انسانی دیجیتال، زبان‌شناسی رایانشی و تحلیل متن گردآوری شده‌اند. همچنین، گزارش‌ها و مستندات فنی مرتبط با نرم‌افزارهای تخصصی مورد استفاده -از جمله: Voyant Tools، Mallet، AntConc و Word2Vec- به‌عنوان منابع تکمیلی مورد بررسی قرار گرفته‌اند.

تحلیل داده‌ها بر پایه روش تحلیل محتوای کیفی، تفسیر نظریات و مقایسه تطبیقی میان رویکردها صورت پذیرفته است. در این راستا، نگارنده با بهره‌گیری از تجربه آموزشی در محیط آموزش غیرحضوری دانشگاه پیام نور، تحلیل‌هایی را از منظر زیسته دانشگاهی و تجربیات عملی در تدریس متون ادبی فارسی در فضای دیجیتال ارائه کرده است.

## کاربست نظریه‌ها و ابزارهای علوم انسانی دیجیتال در تحلیل متون دینی فارسی

کاربرد نظریه خوانش دور (Distant Reading) در تحلیل ادبیات دینی و عرفانی فارسی در عصر دیجیتال، شیوه‌های مطالعه متون ادبی دستخوش تحولاتی چشمگیر شده‌اند. یکی از نظریه‌های بنیادین در این زمینه، نظریه «خوانش دور» (Distant Reading) اثر فرانکو مورتی<sup>۱</sup> (۲۰۱۳) است. این رویکرد با کنار گذاشتن خوانش سنتی متن منفرد، پیشنهاد می‌کند که پژوهشگر با بهره‌گیری از ابزارهای آماری، الگوریتمی و گراف‌محور، به شناسایی الگوهای کلان معنایی، زبانی و سبکی در پیکره‌های بزرگ متنی بپردازد.

۱. فرانکو مورتی (متولد ۱۹۵۰، ایتالیا) استاد ادبیات تطبیقی در دانشگاه استنفورد و بنیان‌گذار Stanford Literary Lab است. او در سال ۲۰۰۰ نخستین بار مفهوم «خوانش دور» (Distant Reading) را معرفی کرد که در مقابل روش سنتی خوانش نزدیک قرار می‌گیرد و بر تحلیل آماری و داده‌محور حجم انبوه متون تأکید دارد. مورتی و آثار برجسته‌ای چون (Distant Reading, 2013) و (Graphs, Maps, Trees 2005) در شکل‌گیری مطالعات ادبیات دیجیتال نقش کلیدی ایفا کرده‌اند.

کاربرد این نظریه در حوزه ادبیات فارسی، به‌ویژه در گونه‌های دینی و عرفانی، می‌تواند افق‌های تازه‌ای بگشاید؛ چراکه با وجود حجم انبوه و ارزشمند متون در این حوزه، روش‌های کلاسیک خوانش، پاسخ‌گوی کاوشی ساختاریافته و همه‌جانبه در این میراث عظیم نیستند.

### ظرفیت‌های قابل تحقق برای ادبیات فارسی

بر اساس نظریهٔ مورتی، می‌توان راهکارهایی عملی برای آیندهٔ پژوهش در ادبیات فارسی پیشنهاد کرد، از جمله:

الف) طراحی پیکره‌های موضوع‌محور پژوهشگران می‌توانند مجموعه‌هایی از متون عرفانی و دینی فارسی را در ژانرهای مختلف گردآوری و سامان‌دهی کنند، از جمله:

غزلیات عرفانی (عطار، عراقی، مولوی، حافظ و...)

مناجات‌نامه‌ها و ادعیهٔ مشهور (مانند مناجات‌نامهٔ خواجه عبدالله انصاری)

تفسیر ادبی-عرفانی قرآن (کشف‌الاسرار، تفسیر ابوالفتوح رازی و...)

این پیکره‌ها باید از نظر زبانی پاک‌سازی<sup>۱</sup>، نشانه‌گذاری و دیجیتالی شوند تا قابلیت تحلیل خودکار با نرم‌افزارهایی چون Voyant Tools، AntConc یا Gensim در محیط پایتون را پیدا کنند.

۱. منظور از پاک‌سازی (Cleaning) فرآیندی است که طی آن متن خام دیجیتالی شده آماده‌ی تحلیل آماری و زبانی می‌شود. این مرحله، پیش‌نیاز حیاتی برای هرگونه تحلیل رایانشی یا زبان‌شناختی است. در مورد متون عرفانی و دینی فارسی، پاک‌سازی شامل چند مرحله‌ی کلیدی است:

مراحل اصلی پاک‌سازی متن:

۱. حذف نویسه‌ها و علائم زائد

- پاک‌کردن علائم غیرزبانی (مثلاً ستاره‌ها، شماره آیات، علائم صفحه‌بندی)

- حذف علائم نگارشی در صورت نیاز (مثلاً نقطه، ویرگول) برای تحلیل آماری ساده‌تر

۲. یکسان‌سازی نویسه‌ها (Normalization)

- تبدیل «ی» عربی به «ی» فارسی، «ک» همزه‌دار) عربی به «ک» فارسی

مثال: «می‌دانم» ← «می‌دانم»

- اصلاح فاصله‌ها: حذف فاصله‌ی اضافی یا اضافه‌کردن فاصله‌های لازم

۳. اصلاح اشتباهات تایپی و اسکن

- اصلاح غلط‌های رایج در نسخه‌های دیجیتالی شده‌ی کهن‌متن‌ها

۴. یک‌دست‌سازی املائی واژگان

ب) تحلیل فراوانی واژگان و مفاهیم کلیدی به جای تفسیر یک اثر منفرد، پژوهشگر می‌تواند بسامد مفاهیم بنیادین را در مجموعه‌ای از آثار استخراج کند. برای نمونه:

در سطح الهی: خدا، نور، اسماء، ذات

در سطح عرفانی: عشق، فنا، وصال، حضور

در سطح اخلاقی: توبه، ورع، زهد، فقر، صبر، توکل، رضا

در سطح بلاغی: استعاره، ایهام، تضاد، تناسب

این تحلیل آماری می‌تواند نشان دهد که کدام مفاهیم در کدام دوره یا نزد کدام نویسنده برجسته‌تر بوده‌اند.

### افق تحلیل روندهای زمانی و سبک‌شناختی

از دیگر ظرفیت‌های نظریه خوانش دور، امکان تحلیل تغییرات مفهومی در بستر زمان است. برای نمونه:

در قرن ششم هجری، تأکید متون بر زهد و تقواست.

در قرن هفتم، مفاهیم عرفان عاشقانه غالب می‌شوند.

در قرن هشتم، زبان بلاغی پیچیده‌تر شده و عناصر ایهام و رمزگرایی افزایش می‌یابد.

این نوع تحلیل، به جای اتکای صرف بر حدس‌های فردی، از داده‌های واقعی برای ترسیم

تحولات سبکی و معنایی بهره می‌گیرد.

### ترسیم شبکه‌های مفهومی

با استخراج هم‌رخدادی واژگان پرتکرار در متون، می‌توان گراف‌هایی از روابط معنایی ترسیم

کرد. برای مثال، مفاهیمی چون «عشق»، «نور»، «دل»، «وصال» و «خدا» معمولاً در کنار هم ظاهر

می‌شوند و کانون‌های معنایی یک جریان فکری را نمایان می‌سازند.

۵. مثلاً انتخاب یک صورت برای واژه‌هایی چون «خویش» / «خویشتن» یا «الله» / «خدا»

حذف یا استانداردسازی علائم نسخه‌پردازان یا مصححان

مانند کروش، برانتز، یا توضیحات و پاورقی‌های اضافه‌شده توسط مصحح

۶. نشانه‌گذاری ساختاری (در صورت نیاز)

۷. جداکردن مصرع‌ها، بندها، آیات یا واحدهای معنایی برای تحلیل موضوعی یا سبک‌شناختی

ضرورت پاک‌سازی از این جهت است که بیشتر ابزارهای تحلیل متنی (مثل LDA، Word2Vec، یا نرم‌افزارهایی چون Voyant و

AntConc) نمی‌توانند با متن‌های ناپالوده یا متناقض از نظر نگارش کار کنند. بدون پاک‌سازی، داده‌های به‌دست‌آمده بی‌دقت یا حتی

گمراه‌کننده خواهند بود.

در ادبیات عرفانی، چنین شبکه‌هایی می‌توانند پیوند مفاهیم عرفانی با ساختار بلاغی متن را آشکار کرده و تفاوت‌های سبکی و معنایی آثار مختلف را به صورت بصری و تحلیلی نشان دهند.

### تحلیل سبکی بر اساس الگوریتم‌های آماری

سبک‌شناسی آماری (Stylometry)، در امتداد نظریه خوانش دور مورثی، امکان تحلیل سبک نویسندگان را بر پایه شاخص‌های کمی فراهم می‌سازد. این شاخص‌ها شامل موارد زیر هستند:

- میانگین طول جملات

- نوع و تنوع واژگان

- توزیع افعال، ضمایر و عناصر دستوری

- ساخت‌های نحوی پرتکرار و الگوهای زبانی خاص

کاربرد این روش‌ها به پژوهشگران اجازه می‌دهد تا متونی را که فاقد مؤلف مشخص‌اند (مانند آثار بی‌نام یا منسوب)، بر اساس سبک زبانی، نسبت‌دهی کنند. همچنین می‌توان تفاوت‌های سبکی میان دو اثر از یک نویسنده را در دوره‌های مختلف زندگی او بررسی کرد و تحول سبک فردی را به صورت داده‌محور تحلیل نمود.

### نمونه تحلیل خلاصه متنی

در ادامه، یک نمونه تحلیل داده‌محور و کلان‌نگر بر اساس نظریه خوانش دور و با بهره‌گیری از ابزارهای متن‌کاوی، از غزل نخست دیوان شمس تبریزی (مولوی، ۱۳۷۸: ۴) ارائه می‌شود:

ای دوست بیا تا غم فردا نخوریم / وین یک دم عمر را غنیمت شمیریم...

در این تحلیل، با استفاده از نرم‌افزارهایی چون Voyant Tools و Word2Vec، شاخص‌هایی نظیر فراوانی واژگان، هم‌رخدادی مفاهیم، ساخت نحوی غالب، و شبکه معنایی واژگان استخراج شده‌اند. نتایج اولیه نشان می‌دهد که مفاهیم «دوستی»، «زمان»، «وصال» و «غنیمت» در این غزل به صورت خوشه‌ای ظاهر شده و ساختار بلاغی آن بر پایه دعوت به حضور و رهایی از آینده‌نگری بنا شده است. همچنین، سبک زبانی مولوی در این غزل با استفاده از شاخص‌های آماری، از جمله بسامد افعال امری، ضمایر مخاطب و ساخت‌های تکرارشونده، قابل شناسایی و مقایسه با دیگر غزلیات او در دوره‌های مختلف است.

ای رَسْتخیزِ ناگهانِ وی رحمتِ بی مُنتها  
 امروز خندان آمدی مِفْتاحِ زندانِ آمدی  
 خورشید را حاجبِ تویی او مید را واجبِ تویی  
 در سینه‌ها برخاسته اندیشه را آراسته  
 خورشید را حاجبِ تویی او مید را واجبِ تویی  
 ما زان دَغَلِ کَژبینِ شده با بی‌گنه در کین شده  
 این سُکر بینِ هِلِ عَقَلِ راوینِ نُقَلِ بینِ هِلِ نُقَلِ را  
 تدبیرِ صَدِ رَنگِ اَفکنی بَرِ رومِ و بَرِ رَنگِ اَفکنی  
 می مال پنهانِ گوشِ جانِ می زه بهانه بَرِ کسان  
 خامش که بس مُسْتَعجَلِمِ رَفتمِ سویی پایِ عَلمِ  
 ای آتشی آفروخته در بیشه‌ اندیشه‌ها  
 بر مُستمدانِ آمدی چون بَخشش و فَضَلِ خدا  
 مطلبِ تویی طالبِ تویی هم مُنتها هم مُبتدا  
 هم خویش حاجتِ خواسته هم خویشانِ کرده روا  
 مطلبِ تویی طالبِ تویی هم مُنتها هم مُبتدا  
 گه مَسْتِ حورالعینِ شده گه مَسْتِ نان و شوربا  
 کز بهرِ نان و بَقَلِ را چندین نشاید ماجرا  
 و اندر میانِ جَنگِ اَفکنی «فی اِصْطِناعِ لا یُری»  
 جان «رَبِّ خَلْصَنِی» زان و الله که لاغ است ای کیا  
 کاغذِ بِنه بِشکن قَلَمِ ساقی دَرآمد الصّلا

### تحلیل واژگانی (Word Frequency Analysis)

با استفاده از نرم‌افزارهایی مانند Voyant Tools یا AntConc، تحلیل برخی واژگان در این

غزل به صورت زیر نمایان می‌شوند:

واژه	بسامد و تکرار	خوشه معنایی
تو / مخاطب	۶ بار	مخاطب قدسی (الهی)
آمدی	۳ بار	ظهور / تجلی
دغل	۲ بار	فریب / انحراف
منتها / مبتدا	۱ بار	مفاهیم فلسفی - عرفانی
نان / شوربا / نقل	۳ بار	تمثیل‌های روزمره / دنیاطلبی

- غلبه با ضمیر دوم شخص «تو» است: نشانه‌ای از گفت‌وگوی عرفانی با حقیقت یا معشوق الهی.

- افعال حرکتی چون «آمدی» فضای تجلی و ظهور دارند.

- واژگان دنیوی (نان، شوربا) در تضاد با مقام قدسی قرار گرفته‌اند (تضاد مفهومی/ بلاغی).

### تحلیل مفهومی (Conceptual Clustering)

تحلیل هم‌رخدادی واژه‌ها نشان می‌دهد که مفاهیم غالب در غزل در سه دسته کلیدی متمرکز هستند: الف. تجلی قدسی: رستخیز ناگهان - رحمت بی‌منتها - روح‌بخش بی‌بدل - لذت علم و عمل. این مفاهیم به حوزه تجلی، رحمت و تعالی معنوی تعلق دارند. ب. فروپاشی عقل جزئی: سُکر / عقل / نقل - تدبیر صد رنگ / جنگ / زنگ. مولانا ساختار عقل ابزاری را در برابر شور الهی قرار می‌دهد. ج. نقد دنیاگرایی: نان / شوربا / حورالعین - دَعْل / کژبین / بی‌گنه در کین. نمایش ناکامی انسان در فهم حقیقت به دلیل درگیر شدن در ظواهر دنیا.

### سبک‌شناسی آماری (Stylometry Snapshot)

در همین یک غزل می‌توان شاخص‌های سبکی زیر را یافت:

ویژگی زبانی	مصادق/نمونه	تفسیر/معنی
تکرار افعال به‌ویژه مخاطب و امری	آمدی / افکنی / ... زنان	ایجاد ریتم و تأکید عاطفی
تقابل مفاهیم	عقل / سُکر، دغل / فضل / شوربا / علم	از اصول بلاغت زبان عرفانی
فراوانی ضمائر (دوم شخص)	ای... / تویی / تو	گفتگوی رودرو و صریح با معشوق الهی
استعاره‌های فرارونده	آتش در بیشه اندیشه / حاجب / خورشید / کلید زندان	استعاره از حلول و نفوذ حقیقت در دل عارف

### نقشه شبکه‌ای مفاهیم (Semantic Network Map)

با تحلیل وابستگی معنایی، می‌توان گرافی از پیوند واژه‌ها ساخت:

مرکز شبکه: «تو»

در خطاب به ضمیر «تو» شعاع‌هایی به مفاهیم زیر منشعب می‌شوند:

— آتش ← بیشه اندیشه ← دگرگونی معرفتی

- خورشید ← حاجب ← روشنی و روشننگری
  - مفتاح زندان ← رهایی از محدودیت
  - رحمت، بخشش، فضل ← لطف اله
  - علم و عمل ← غایت سلوک
- در حاشیه شبکه: مفاهیم منفی: (دغل، شوربا، حورالعین ← دنیاگرایی، فریب)، نشان می‌دهد که غزل ساختاری قطبی شده دارد: (حقیقت / دنیا)؛ و تمام استعاره‌ها در خدمت همین دو قطب هستند.

### جمع بندی نمونه تحلیل

این غزل مولانا با محوریت «ظهور حقیقت در هستی»، نمونه‌ای از متنی است که ساختار آن برای تحلیل داده محور بسیار مناسب است. ابزارهای دیجیتال در این زمینه می‌توانند فراتر از تأویل‌های فردی، چشم‌اندازی کل‌نگر از روابط مفاهیم، سبک‌ها، و سلوک فکری مولانا فراهم کنند. کاربرد در تحلیل‌های گسترده‌تر (Macro-Level Utility) اگر چنین تحلیلی بر صدها غزل عرفانی اعمال شود، می‌توان:

- بسامد مضامین را در آثار مختلف بررسی کرد.
- روند تحول عرفانی مولوی را از آغاز تا پایان شناسایی کرد.
- سبک‌شناسی تطبیقی بین عراقی و عطار و حافظ و... ارائه داد.
- شبکه‌های مفهومی مشترک یا منحصر به فرد هر شاعر را ترسیم نمود.

### ترکیب زبان‌شناسی پیکره‌ای و تحلیل ادبی در بستر علوم انسانی دیجیتال

در این بخش از مقاله به یکی از جدیدترین و تحول‌آفرین‌ترین رویکردها در مطالعات ادبی یعنی ترکیب زبان‌شناسی پیکره‌ای (Corpus Linguistics) و مطالعات ادبی (Literary Studies) پرداخته می‌شود. این حوزه که بخشی از مطالعات ادبی دیجیتال (Digital Literary Studies) به‌شمار می‌آید به پژوهشگران امکان می‌دهد تا به‌جای تحلیل سنتی مبتنی بر نمونه‌های محدود بتوانند هزاران یا حتی میلیون‌ها واژه را با کمک ابزارهای محاسباتی بررسی کنند. در این‌جا به تفصیل به چیستی، نظریه پردازان، ابزارها، مزایا، چالش‌ها و کاربردهای خاص آن در ادبیات فارسی (به‌ویژه شعر دینی و عرفانی) خواهیم پرداخت.

## زبان‌شناسی پیکره‌ای در مطالعات ادبی

زبان‌شناسی پیکره‌ای شاخه‌ای از زبان‌شناسی است که بر تحلیل مجموعه‌های بزرگ متنی (corpora) تمرکز دارد. این مجموعه‌ها اغلب به صورت دیجیتالی فراهم می‌شوند و به کمک ابزارهای تحلیل رایانه‌ای پردازش می‌گردند. وقتی این رویکرد وارد حوزه ادبیات می‌شود، تمرکز از خواندن تفسیری (close reading) به تحلیل کلان (distant reading) تغییر می‌کند.

### نمایندگان اصلی این حوزه

سه تن از چهره‌های شاخص این رویکرد عبارت‌اند از:

- متیو جوکرز<sup>۱</sup> نویسنده کتاب تحلیل کلان‌نگر<sup>۲</sup> (2013) از نخستین کسانی است که تحلیل‌های کلان‌نگر را بر پیکره‌های بزرگ داستانی و رمان‌های قرن نوزدهم به کار برد.
- آندرو پِیپر<sup>۳</sup> استاد دانشگاه مک‌گیل که آثار متعددی درباره ادبیات و داده‌کاوی ادبی منتشر کرده و بر سبک‌شناسی رایانشی تمرکز دارد.
- تد آندروود<sup>۴</sup>: پژوهشگری که با تحلیل سبک‌های روایی در «پیکره‌های چندقرنی» به بازخوانی تاریخ ادبیات از طریق ماشین پرداخته است.

### شکل کاربست در متن

هدف از ترکیب زبان‌شناسی پیکره‌ای با مطالعات ادبی، ساخت و تحلیل یک پیکره دیجیتالی از اشعار دینی فارسی است؛ پیکره‌ای شامل متونی چون مناجات‌نامه‌ها، اشعار عرفانی، ادعیه منظوم و آثار شعرایی مانند سنایی، عطار، مولوی، خواجه عبدالله انصاری و... این پیکره با ابزارهای متن‌کاوی پردازش می‌شود تا الگوهای زبانی، فراوانی مفاهیم الهی، ساخت‌های نحوی دعایی و کاربرد واژگان احساسی در آن بررسی شود.

- 
1. Matthew Jockers
  2. *Macroanalysis*
  3. Andrew Piper
  4. Ted Underwood

برای مثال می‌توان پرسید:

- فراوان‌ترین صفات خدا در شعر مناجاتی چیست؟
- چه تفاوت‌هایی در ساختارهای نحوی میان مناجات‌های سنایی و مولوی وجود دارد؟
- واژگان مرتبط با عشق، خوف، رجاء و فنا در هر دوره یا در آثار هر شاعر چه بسامدی دارند؟

### معرفی ابزارهای تحلیل متن دیجیتال

پژوهشگر در این رویکرد با گردآوری متونی بزرگ از آثار ادبی و واردکردن آن‌ها به نرم‌افزارهای تحلیل متن، به استخراج الگوهای زبانی، سبکی، و معنایی می‌پردازد. برخی از ابزارهای پرکاربرد عبارت‌اند از:

۱. «Voyant Tools»: ابزار رایگان و آنلاین برای تحلیل‌های مقدماتی مانند شمارش واژگان، هم‌رخدادی، ترسیم نمودارهای زمانی و...
۲. «AntConc»: نرم‌افزاری حرفه‌ای برای تحلیل واژه‌های هم‌نشین، واژه‌های کلیدی، الگوهای سبکی و تحلیل‌های آماری پیچیده‌تر.
۳. «NLTK (Natural Language Toolkit)»: مجموعه‌ای از ابزارهای زبان‌پردازی در زبان برنامه‌نویسی پایتون، برای تحلیل‌های پیشرفته‌تر مانند تحلیل نحوی، معنایی و حتی یادگیری ماشین در زبان.
۴. «Voyant Tools»: نرم‌افزاری آنلاین و منبع‌باز جهت تحلیل متن است که توسط Stéfan Sinclair و Geoffrey Rockwell طراحی شده و به‌طور گسترده در مطالعات علوم انسانی دیجیتال برای کاوش در متون ادبی، تاریخی و اجتماعی به‌کار می‌رود. این نرم‌افزار قابلیت‌هایی چون: شمارش واژگان (Word Frequency)؛ شناسایی کلیدواژه‌ها (Keywords in Context - KWIC)؛ ترسیم هم‌زمانی واژگان (Co-occurrence)؛ رسم نمودارهای تغییر واژگان در طول متن (Trends)؛ شبکه‌سازی مفاهیم را فراهم می‌آورد. این ابزار برای مطالعات تطبیقی متون و بررسی زبانی آثار کلاسیک فارسی نیز قابل استفاده است. مزیت آن سادگی رابط کاربری، پشتیبانی از زبان فارسی و عدم نیاز به نصب است.

۵. «MALLET (MACHINE Learning for Language Toolkit)»: یک بسته پیشرفته و آفلاین تحلیل متنی بر پایه زبان جاوا Java است که توسط دانشگاه ماساچوست توسعه یافته و برای پژوهش‌های عمیق در زمینه‌های علوم انسانی، زبان‌شناسی محاسباتی و داده‌کاوی متنی کاربرد دارد. مهم‌ترین قابلیت‌های نرم‌افزار:

- مدل‌سازی موضوعی (Topic Modeling) با استفاده از الگوریتم‌های Latent (LDA) Dirichlet Allocation)؛

- دسته‌بندی متون (Text Classification)؛

- تحلیل رگرسیون<sup>۱</sup> متنی؛

- تشخیص الگوهای پنهان در حجم انبوهی از متن‌ها؛

این نرم‌افزار به دلیل انعطاف‌پذیری بالا و دقت الگوریتم‌های آماری، برای تحلیل متون فارسی (در صورت پیش‌پردازش مناسب) نیز قابل استفاده است؛ هر چند نیاز به تسلط فنی بیشتری نسبت به Voyant دارد.

۱. «رگرسیون» (Regression) اصطلاح آماری و تحلیلی است که در زبان فارسی معمولاً با یکی از معادل‌های زیر ترجمه می‌شود: معادل‌های فارسی:

۱. بازگشت: رایج‌ترین ترجمه، به‌ویژه در متون آماری و تحلیلی.

مثال: تحلیل رگرسیون = تحلیل بازگشت

۲. پس‌روی: در برخی متون تخصصی‌تر به‌ویژه در روان‌شناسی.

۳. برگشت‌پذیری یا واگرایی و همگرایی - در برخی کاربردهای خاص، ولی کم‌کاربردتر.

درواقع رگرسیون روشی آماری است برای بررسی رابطه بین یک متغیر وابسته (مثلاً نمره یا درآمد) و یک یا چند متغیر مستقل (مثلاً سن، تحصیلات، جنسیت و...)

مثلاً می‌خواهیم بفهمیم آیا بین ساعت مطالعه و نمره‌ی امتحان رابطه‌ای وجود دارد یا نه.

- متغیر مستقل: ساعت مطالعه

- متغیر وابسته: نمره امتحان

با رگرسیون می‌توانیم بفهمیم:

- آیا با افزایش ساعت مطالعه، نمره هم بالا می‌رود یا نه؟

- چقدر از تغییرات نمره را می‌توان با ساعت مطالعه توضیح داد؟

تحلیل رگرسیون متنی چیست؟

«تحلیل رگرسیون متنی» یعنی استفاده از رگرسیون برای تحلیل رابطه‌ی بین ویژگی‌های متنی (مثل تعداد واژه‌ها، طول جمله‌ها، بسامد واژگان خاص، سبک نوشتار، تکرار مفاهیم) و یک متغیر هدف (مثلاً رضایت مخاطب، پیچیدگی متن، سطح خوانایی یا موفقیت در جذب مخاطب). نمونه‌ی کاربرد:

در مطالعات زبان‌شناسی رایانشی (computational linguistics) یا زبان‌شناسی کاربردی، ممکن است بررسی شود:

- آیا تعداد فعل‌های مجهول در یک مقاله بر درک‌پذیری آن اثر دارد؟

- یا آیا طول جملات پیش‌بینی‌کننده‌ی خوبی برای سطح خوانایی متن هستند؟

برای انجام مدل‌سازی موضوعی و تحلیل شبکه معنایی، ابزارهای نرم‌افزاری متعددی وجود دارد: «MALLET (MACHINE Learning for Language Toolkit)»: ابزار خط فرمانی که برای پیاده‌سازی الگوریتم LDA بسیار مناسب و پرکاربرد است.

«Gensim در پایتون»: کتابخانه‌ای بسیار قدرتمند برای مدل‌سازی موضوعی که علاوه بر LDA، مدل‌های دیگری مانند LSI، HDP و Word2Vec را نیز پیاده‌سازی می‌کند.

«Voyant Tools»: هرچند بیشتر برای تحلیل‌های مقدماتی متن کاربرد دارد، ولی قابلیت‌های اولیه‌ای برای مدل‌سازی معنایی نیز ارائه می‌دهد.

«Gephi»: نرم‌افزار متن‌باز برای ترسیم گراف‌های مفهومی و تحلیل شبکه‌ای (network analysis) مفاهیم. خروجی مدل‌های معنایی را می‌توان در آن به صورت نقشه‌های مفهومی ترسیم کرد.

### مزایا و امکانات روش

۱. پرهیز از سوگیری فردی: برخلاف خوانش سنتی، در این رویکرد داده‌ها به صورت عینی و کمی تحلیل می‌شوند.
۲. بررسی کلان متون: به جای تحلیل چند بیت یا چند صفحه، می‌توان الگوهای کلان در صدها یا هزاران بیت را استخراج کرد.
۳. امکان مقایسه دقیق سبک شاعران: با تحلیل آماری می‌توان سبک شعرا را از منظر نحوی، واژگانی و معنایی مقایسه کرد.
۴. قابلیت ترکیب با تحلیل‌های معنایی یا بلاغی: می‌توان داده‌های پیکره را با تحلیل استعاره، نماد یا انواع بلاغی ترکیب کرد.

### چالش‌ها و محدودیت‌ها

با آن‌که بسیاری از بسترها و نرم‌افزارهای تحلیل متنی همچون AntConc و Lexos، Voyant Tools در اصل برای متون انگلیسی طراحی شده‌اند، اما در سال‌های اخیر قابلیت استفاده نسبی آن‌ها برای متون فارسی نیز مورد توجه پژوهشگران قرار گرفته است. البته باید توجه داشت که برخی از این ابزارها به دلیل ویژگی‌های خاص زبان فارسی همچون راست‌به‌چپ بودن نوشتار، پیوستگی و جداسدگی ضمایر و

پیچیدگی‌های صرفی و نحوی، عملکردی محدودتر دارند. با این حال، تجربه نشان داده است که با اصلاحات اولیه در پیش‌پردازش متن می‌توان از برخی قابلیت‌های آماری و گرافیکی این ابزارها در مطالعات فارسی نیز بهره گرفت. از این رو به برخی از این محدودیت‌ها اشاره می‌شود:

۱. مشکلات متن‌پردازی فارسی: زبان فارسی از نظر صرفی و نحوی با زبان‌های غربی تفاوت دارد و اغلب ابزارها برای انگلیسی طراحی شده‌اند.
۲. دشواری تصحیح متون: بسیاری از اشعار فارسی نسخه‌برداری یا تایپ‌شده‌اند و دارای خطاهای املایی یا چاپی هستند.
۳. نیاز به دانش فنی: استفاده پیشرفته از ابزارهایی چون NLTK نیازمند آشنایی با برنامه‌نویسی است.
۴. ابهام در تفسیر داده‌ها: داده‌های عددی باید با نگاه بلاغی و معنایی همراه شوند تا معنای دقیق ادبی پیدا کنند.

باری، ترکیب زبان‌شناسی پیکره‌ای و مطالعات ادبی افق جدیدی در پژوهش‌های شعر دینی فارسی می‌گشاید. این روش به ما اجازه می‌دهد تا با نگاهی جامع و علمی، ساختار معنایی و زبانی اشعار عرفانی و مناجاتی را بازخوانی کنیم. از این رهگذر، نه تنها می‌توان سیر تحول زبان دعا و نیایش در ادب فارسی را بررسی کرد، بلکه می‌توان گفت‌وگوی عارف با خداوند را در لایه‌های مختلف زبانی و معنایی تحلیل کرد؛ تحلیلی که به تنهایی با روش‌های سنتی ممکن نبود.

### مدل‌سازی موضوعی و تحلیل شبکه معنایی

با گسترش مطالعات ادبی دیجیتال، تحلیل‌های سنتی متون که اغلب به خوانش تفسیری محدود بودند جای خود را به روش‌هایی داده‌اند که به کمک الگوریتم‌های آماری، الگوهای پنهان در متون بزرگ را کشف می‌کنند. یکی از مهم‌ترین این روش‌ها Topic Modeling یا مدل‌سازی موضوعی است؛ روشی که به پژوهشگر امکان می‌دهد بدون نیاز به برچسب‌گذاری دستی، خوشه‌هایی از مفاهیم و موضوعات پنهان را در مجموعه‌ای از متون شناسایی کند.

مدل‌سازی موضوعی یک روش یادگیری ماشین (Machine Learning) و بدون برچسب‌گذاری و نظارت از بالا (Unsupervised) است که در آن، موضوعات (Topics) به‌عنوان مجموعه‌ای از واژگان پرتکرار که اغلب با یکدیگر هم‌رخداد دارند، شناسایی می‌شوند. شناخته‌شده‌ترین الگوریتم این حوزه، Latent Dirichlet Allocation (LDA) است که توسط Blei, Ng و Jordan در سال ۲۰۰۳ معرفی شد.

در عمل، LDA هر متن را ترکیبی از چند موضوع می‌بیند و هر موضوع را نیز ترکیبی آماری از چند واژه. برای مثال، اگر موضوعی با واژه‌های «نور»، «طلوع»، «خورشید»، «روشنایی»، و «سپیده» شناخته شود، می‌توان آن را به‌صورت شهودی «موضوع تجلی» یا «نور الهی» نام‌گذاری کرد.

### پیشینه دانشگاهی و کاربردهای جهانی

مدل‌سازی موضوعی و تحلیل شبکه معنایی اکنون به یکی از ارکان مطالعات انسانی دیجیتال تبدیل شده‌اند. از برجسته‌ترین پروژه‌های دانشگاهی در این زمینه می‌توان به موارد زیر اشاره کرد: «Stanford Literary Lab»: که در آن LDA برای تحلیل موضوعات در رمان‌های قرن نوزدهم و بیستم به کار رفته است.

«Mining the Dispatch (University of Richmond)»: پروژه‌ای که موضوعات پنهان در روزنامه‌های قرن نوزدهم آمریکا را بازبایی و تحلیل می‌کند. پژوهش‌هایی در دانشگاه‌های ام‌آی‌تی، هاروارد و کمبریج از LDA برای بررسی سیر تحول مضامین فلسفی، علمی یا دینی در متون بزرگ استفاده کرده‌اند.

### کاربرد در ادبیات فارسی اسلامی

مفاهیم در ادبیات فارسی به‌ویژه مفاهیم عرفانی و دینی اغلب به‌شکل نمادین و بینامتنی (intertextual) حضور دارند. همین امر موجب می‌شود که تحلیل سنتی نتواند درکی منسجم از کل مجموعه بدهد. مدل‌سازی موضوعی در این حوزه می‌تواند تحول‌آفرین باشد. برای مثال:

۱. منظور این است که هیچ برچسب یا پاسخ مشخصی به الگوریتم داده نمی‌شود و خود سیستم باید الگوها و ساختارها را در داده‌ها کشف کند (مثلاً خوشه‌بندی یا مدل‌سازی موضوعی). مثلاً در مدل‌سازی موضوعی (مثل LDA)، داده‌ها فقط متون خام هستند؛ الگوریتم خودش «موضوعات» را پیدا می‌کند.

۱. استخراج موضوعات کلیدی در اشعار عرفانی: تحلیل پیکره‌ای از دیوان مولوی یا عطار می‌تواند به استخراج خودکار موضوعاتی مانند «نور»، «عشق»، «ترس از خدا»، «توبه»، «وصال»، «فنا»، «وصل»، «غیبت» و غیره بینجامد. پژوهشگر می‌تواند نسبت حضور هر یک از این موضوعات در دوره‌های مختلف زندگی شاعر یا سبک‌های مختلف شعری (غزل، مثنوی، رباعی) را نیز تحلیل کند.
۲. ترسیم شبکه معنایی مفاهیم: با استفاده از ابزار Gephi یا Pajek، می‌توان شبکه‌ای از مفاهیم ایجاد کرد که نشان دهد چگونه واژه‌های خاص (مثلاً «نور») با دیگر واژه‌ها (مثلاً «دل»، «وصل»، «ظهور») در یک ساختار معنایی پیوند دارند. این امر به درک نظام اندیشه عرفانی در زبان کمک می‌کند.
۳. مقایسه میان شاعران: مدل LDA می‌تواند بر روی آثار چند شاعر مانند سنایی، مولوی و عطار پیاده شود و با رسم نمودارهای پراکنش موضوعات، نشان داد که مثلاً مولوی بیشتر بر «وصل» و «نور» تأکید دارد، در حالی که سنایی بر «خوف»، «عدالت» متمرکز است.
۴. تحلیل تفاسیر ادبی دینی: متونی چون کشف‌الاسرار و فیه مافیه منابعی بسیار غنی برای تحلیل معنایی هستند. مدل‌سازی موضوعی می‌تواند مضامین تکرارشونده در تفسیرهای آن‌ها را بازیابی کند، مثلاً مضامین «عشق»، «شهود»، «مقام قرب» و غیره.

### نقاط قوت این روش

- کشف الگوهای پنهان معنایی بدون دخالت پیش‌فرض پژوهشگر؛
- قابل استفاده برای متون بسیار طولانی و متنوع؛
- قابلیت تصویری‌سازی (Visualization) با گراف‌های مفهومی؛
- امکان تحلیل تغییر موضوعات در طول زمان یا در آثار مختلف.

### چالش‌ها و محدودیت‌ها

نیاز به پیکره پاک و دیجیتالی شده (cleaned corpus): مدل LDA به شدت به کیفیت داده ورودی حساس است.

پیچیدگی در تفسیر موضوعات آماری: خروجی LDA صرفاً لیستی از واژگان است که نیاز به تفسیر انسانی دارد.

- نبود واژه‌نامه معنایی دقیق برای فارسی کلاسیک: واژگان عرفانی گاه معانی متعددی دارند که الگوریتم به تنهایی نمی‌تواند تمییز دهد.

- چالش پردازش زبان فارسی: وجود پسوندها، شناسه‌ها و صرف پیچیده در فارسی کلاسیک، نیازمند پیش‌پردازش حرفه‌ای است.

در کل مدل‌سازی موضوعی (به‌ویژه LDA) و تحلیل شبکه معنایی، دریچه‌ای نو به تحلیل متون دینی و عرفانی فارسی می‌گشاید. این رویکرد به جای اتکا به تحلیل‌های ذهنی یا نمونه‌های محدود، امکان بررسی صدها هزار واژه را در قالب الگوریتمی علمی فراهم می‌سازد. پژوهشگر برای نخستین بار می‌تواند شبکه موضوعات، خوشه‌های معنایی و سیر تحول مفاهیم را نه صرفاً بر مبنای شهود، بلکه با تکیه بر داده و آمار استخراج کند.

این روش در چشم‌اندازی بلندمدت می‌تواند پایه‌ای برای تدوین نقشه معنایی عرفان فارسی و حتی تبارشناسی (genealogy) مفاهیم الهیاتی در شعر فارسی باشد. چنین امری می‌تواند جایگاه ادبیات فارسی را در مطالعات تطبیقی جهانی بازتعریف کند.

### سبک‌شناسی آماری (Stylometry)

در کنار معناشناسی و مدل‌سازی موضوعی یکی از روش‌های نیرومند در حوزه ادبیات دیجیتال، سبک‌شناسی آماری (Stylometry) است. این حوزه از تحلیل متن به جای تمرکز بر محتوا یا مضمون، ویژگی‌های کمی نوشتار را در مرکز توجه قرار می‌دهد: بسامد واژگان، ساختار نحوی، طول جملات، واژگان پرکاربرد و حتی علائم نگارشی. این روش به پژوهشگر اجازه می‌دهد تا سبک نویسنده، دوره تاریخی، یا ژانر را بر مبنای داده‌های آماری بازشناسد.

## پیشینه علمی و چهره‌های برجسته

یکی از بنیان‌گذاران سبک‌شناسی آماری مدرن، جان بارو است. وی در دهه ۱۹۹۰ با معرفی شیوه Delta، امکان تشخیص نویسنده متون ناشناس بر اساس بسامد واژه‌های پرکاربرد (function words) را فراهم کرد. Delta میزان شباهت سبک دو متن را بر حسب فاصله آماری محاسبه می‌کند و می‌تواند نشان دهد که یک متن بیشتر به سبک کدام نویسنده نزدیک است.

در سال‌های اخیر، پروژه Stylometry with R در دانشگاه‌های اروپای مرکزی (به‌ویژه لهستان و اتریش) این رویکرد را با کمک نرم‌افزار R گسترش داده و برای تحلیل سبک در صدها هزار سند تاریخی، نامه، شعر و مقاله به کار گرفته است.

Stylometry بر این فرض استوار است که هر نویسنده، سبکی آماری خاص خود دارد حتی زمانی که آگاهانه سعی در تقلید یا پنهان‌کاری دارد. این سبک در موارد زیر نمود پیدا می‌کند:

- بسامد واژه‌های پرکاربرد (مثلاً «و»، «که»، «از»، «در» و...)

- طول و پیچیدگی جملات

- استفاده از واژگان خاص یا نادر

- الگوهای نحوی (مثلاً نوع ترجیحات فعل، ساختار وصفی)

- بسامد حروف یا صامت‌ها و مصوت‌ها

حتی دو متنی که از نظر موضوعی مشابه‌اند (مثلاً دو مناجات‌نامه) می‌توانند از نظر سبک نگارشی تفاوت‌های فاحشی داشته باشند. این تفاوت‌ها را می‌توان با ابزارهای آماری تحلیل و مدل‌سازی کرد.

## ابزارهای فنی

برای اجرای سبک‌شناسی آماری، ابزارهای متن‌باز متعددی در دسترس‌اند:

stylo در محیط R: بسته نرم‌افزاری بسیار کامل که تحلیل متنی را با تکنیک‌هایی مانند PCA

تحلیل مؤلفه‌های اصلی Cluster Analysis، و متد Delta انجام می‌دهد.

«JGAAP (Java Graphical Authorship Attribution Program)»: نرم‌افزاری

کاربرپسند برای تحلیل سبک با تمرکز بر تشخیص نویسنده.

«Signature Stylometric Suitek»: مجموعه‌ای از ابزارهای تحلیلی برای سبک‌شناسی در

متون ادبی و تاریخی.

## کاربرد در ادبیات

سبک‌شناسی آماری، به‌ویژه در ادبیات دینی و عرفانی فارسی، فرصت‌هایی بی‌سابقه برای شناسایی نویسندگان، تمایزگذاری سبکی، و تحلیل‌های دوره‌ای فراهم می‌کند. در ادامه چند نمونه کاربردی آورده شده است:

تشخیص سبک و نویسنده در متون بی‌نشان: یکی از چالش‌های جدی در متون دینی فارسی، عدم انتساب قطعی برخی متون به نویسنده‌ای خاص است. به‌عنوان مثال، برخی مناجات‌نامه‌ها یا خطبه‌های دینی یا اشعار عربی و فارسی به چهره‌هایی مانند امام علی (ع)، خواجه عبدالله انصاری یا عین‌القضات نسبت داده شده‌اند، اما سندیت تاریخی روشنی ندارند. یا گاهی نام نویسنده نسخه خطی از صفحات نخست کتاب افتاده است و صاحب اثر مشخص نیست. سبک‌شناسی آماری با مقایسه سبک نوشتاری این متون با آثار مسلم آن نویسندگان می‌تواند نسبت‌های احتمالی را ارزیابی کند.

تحلیل سبک‌های عرفانی در دوره‌های تاریخی: با ایجاد پیکره‌هایی از متون عرفانی در دوره‌های مختلف (مثلاً دوره غزنوی، سلجوقی، ایلخانی)، می‌توان به بررسی تفاوت‌های سبکی پرداخت. برای مثال:

- آیا نویسندگان سلجوقی تمایل بیشتری به استعاره‌های عرفانی داشتند؟

- آیا طول جملات در دوره مغول کاهش یافته است؟

- بسامد استفاده از واژه‌های عربی چگونه در سبک نگارش تغییر کرده است؟

تحلیل تغییرات سبکی در آثار یک نویسنده: یکی از فرصت‌های قابل توجه، تحلیل تحول سبکی در

آثار یک نویسنده بزرگ است. مثلاً در آثار مولوی:

- در مثنوی معنوی، زبان عمدتاً روایی، تمثیلی و آموزش‌محور است؛

- در غزلیات شمس، زبان شاعرانه، ایهام‌دار و موسیقایی‌تر می‌شود.

Stylometry می‌تواند با تحلیل کمی بسامد واژگان، طول مصراع‌ها، یا استفاده از واژگان عربی و فارسی، این تحول را مستند کند.

تحلیل بلاغت در خطبه‌ها یا دعاها: متونی چون دعای عرفه یا خطبه شمشقیه، به واسطه بلاغت خاص خود شناخته می‌شوند. سبک‌شناسی آماری می‌تواند بسامد استفاده از قالب‌های نحوی خاص، واژگان ردیف‌شونده، یا ساختارهای جناس‌گونه را در آن‌ها تحلیل کرده و تفاوت آن‌ها را با متون غیرادبی نشان دهد.

### نقاط قوت و محدودیت‌ها

نقاط قوت عبارت‌اند از:

- مبتنی بر داده‌های عینی و آماری؛ دور از تفسیرهای ذهنی
  - مناسب برای تحلیل متون بزرگ و مقایسه‌ای
  - قابل استفاده برای پژوهش‌های میان‌رشته‌ای (ادبیات، تاریخ، زبان‌شناسی، رایانه)
  - محدودیت‌ها عبارت‌اند از:
  - حساس به کیفیت پیش‌پردازش (مثلاً حذف صحیح شناسه‌ها در فارسی)
  - لزوم تهیه پیکره‌های تمیز و همگن
  - عدم قابلیت تفسیر معنایی یعنی نمی‌گوید چه گفته شده، بلکه چگونه گفته شده.
- در کل Stylometry یا سبک‌شناسی آماری، ابزاری مهم در تحلیل ادبیات فارسی اسلامی دیجیتال است. این رویکرد به پژوهشگر اجازه می‌دهد تا از محدوده تفسیر ذهنی فراتر رود و با داده‌های آماری، سبک و لحن نویسنده را تحلیل کند. در متونی مانند اشعار عرفانی، دعاها، تفاسیر و مناجات‌نامه‌ها، سبک نگارش نه تنها نقش بلاغی، بلکه نقش معنایی و عرفانی نیز دارد؛ از این رو، تحلیل سبک می‌تواند بخشی از پیام متون را رمزگشایی کند. در کنار روش‌هایی مانند مدل‌سازی موضوعی، تحلیل شبکه معنایی و مطالعات پیکره‌ای، Stylometry نقشی مهم در شکل‌گیری آینده «ادبیات‌شناسی محاسباتی فارسی» ایفا خواهد کرد.

## نتیجه‌گیری

در عصر حاضر که زیست‌بوم فرهنگی و ادبیات فارسی به شکل فزاینده‌ای تحت تأثیر فناوری‌های دیجیتال قرار گرفته است، نیاز به روش‌ها و چارچوب‌های نوین پژوهشی بیش از هر زمان دیگری احساس می‌شود. این مقاله با مرور و معرفی چهار رویکرد برجسته و هم‌افزا در تحلیل متون ادبی - شامل خوانش دور (Distant Reading)، زبان‌شناسی پیکره‌ای (Corpus Linguistics)، مدل‌سازی موضوعی (Topic Modeling) و سبک‌شناسی آماری (Stylometry) - نشان داد که چگونه می‌توان به کمک ابزارهای دیجیتال، حجم گسترده‌ای از متون عرفانی و دینی فارسی را به شیوه‌ای ساختاریافته، کمی و عمیق تحلیل کرد.

این رویکردهای ترکیبی، علاوه بر تسهیل کشف الگوهای معنایی پنهان، امکان شناسایی تغییرات سبک و گفتمان را در بازه‌های تاریخی مختلف فراهم می‌آورند. به‌ویژه در حوزه ادبیات دینی و عرفانی فارسی که پیش‌تر بیشتر متکی بر تحلیل‌های فردی و تفسیرهای سنتی بود، این فناوری‌ها افق‌های تازه‌ای به روی پژوهشگران گشوده‌اند.

افزون بر این، استفاده از این روش‌ها در زیست‌بوم دیجیتال به پژوهشگران امکان می‌دهد تا چالش‌های اساسی نظیر ناپایداری موضوعات ادبی، تداخل ژانرها و تغییرات زبانی را بهتر درک کرده و به راهکارهای نوآورانه برای حفظ و بازآفرینی زبان و ادبیات فارسی دست یابند. به عبارتی، این رویکردها هم ابزار تحلیلی و هم راهبردی برای ارتقای کیفیت و عمق مطالعات ادبی فراهم می‌کنند. در نهایت، این مقاله تأکید دارد که ادبیات فارسی دیجیتال افزون بر این‌که زمینه‌ای برای حفظ میراث کهن است، عرصه‌ای پویا برای نوآوری‌های پژوهشی و ارتباط میان‌رشته‌ای نیز محسوب می‌شود که با بهره‌گیری از فناوری‌های داده‌محور می‌تواند به رشد کیفی و کمی مطالعات ادبی فارسی کمک شایانی کند.

## پیشنهادهایی برای پژوهش‌های آتی و نکات کاربردی

۱. گسترش ساخت پیکره‌های دیجیتال تخصصی و بومی: ایجاد و توسعه پیکره‌های متنی بزرگ و منسجم در حوزه‌های مختلف ادبیات فارسی، به‌ویژه ادبیات عرفانی و دینی با رعایت استانداردهای کدگذاری و نشانه‌گذاری زبانی. این امر زیرساخت محکمی برای تحلیل‌های کمی و داده‌محور فراهم می‌کند.

۲. توسعه ابزارهای تحلیل زبان فارسی: ساخت و بومی‌سازی نرم‌افزارها و ابزارهای تحلیل متن، شامل پردازش زبان طبیعی (NLP) مخصوص زبان فارسی، که بتوانند ویژگی‌های ساختاری و معنایی خاص زبان فارسی را به‌درستی شناسایی و تحلیل کنند.
۳. تحلیل تطبیقی روندهای معنایی و سبکی: انجام مطالعات تطبیقی میان دوره‌های مختلف تاریخی و ژانرهای متنوع ادبیات فارسی با استفاده از تکنیک‌های سبک‌شناسی آماری و مدل‌سازی موضوعی برای فهم بهتر روندهای تغییر و تحولات زبانی و مفهومی.
۴. کاربرد مدل‌های embedding و هوش مصنوعی در تحلیل ادبیات: بسط استفاده از مدل‌های بردار معنایی (BERT، Word2Vec) و غیره) در تحلیل مفاهیم پیچیده عرفانی و دینی، به منظور شناسایی لایه‌های پنهان معنایی و روابط متقابل میان واژگان و مفاهیم.
۵. مطالعات میان‌رشته‌ای و همکاری‌های علمی: تشویق به پژوهش‌های میان‌رشته‌ای که در آن زبان‌شناسی، ادبیات، علوم داده و علوم کامپیوتر با هم تلفیق شوند تا روش‌های نوین و کارآمدتری برای تحلیل متون ارائه شود.
۶. تحلیل داده‌های کاربرمحور در فضای دیجیتال: مطالعه تعامل کاربران و مخاطبان با متون ادبی در شبکه‌های اجتماعی و پلتفرم‌های دیجیتال برای بررسی تأثیرات زیست‌بوم دیجیتال بر شیوه‌های تولید و مصرف ادبیات.
۷. توسعه نظام‌های اعتبارسنجی و آرشیو دیجیتال: ایجاد سامانه‌های مدرن برای ارزیابی کیفیت آثار ادبی دیجیتال، بایگانی منسجم و دسترسی آسان به منابع ادبی دیجیتال با حفظ حقوق مولفان و تضمین اصالت آثار.
۸. پژوهش‌های آموزشی و کاربردی: بررسی کاربرد روش‌های تحلیل داده‌محور در آموزش زبان و ادبیات فارسی، به ویژه در آموزش از راه دور، برای بهبود فهم و درک دانشجویان از متون ادبی پیچیده.

## فهرست نرم افزارها و نشانی وبسایت های مربوط

۱. Voyant Tools، آدرس وبسایت: <https://voyant-tools.org>، (یک ابزار تحت وب رایگان برای تحلیل متون و زبان شناسی پیکره ای که امکان تحلیل آماری و بصری سازی داده های متنی را به صورت تعاملی فراهم می کند).
۲. MALLET (MACHINE Learning for Language Toolkit)، آدرس وبسایت: <http://mallet.cs.umass.edu>، (یک بسته نرم افزاری متن باز برای یادگیری ماشینی، به ویژه در حوزه مدل سازی موضوعی (Topic Modeling) و تحلیل متون).
۳. AntConc، آدرس وبسایت: <https://www.laurenceanthony.net/software/antconc>، (نرم افزاری رایگان برای تحلیل پیکره متنی و جستجوی عبارات، آماره گیری فراوانی لغات و دیگر تحلیل های ساده متنی).
۴. Gensim، آدرس وبسایت: <https://radimrehurek.com/gensim>، (کتابخانه ای متن باز در زبان برنامه نویسی پایتون برای تحلیل موضوعی، مدل سازی بردارهای کلمات (Word Embeddings) و یادگیری معنایی).
۵. Gephi، آدرس وبسایت: <https://gephi.org>، (نرم افزاری آزاد و قدرتمند برای تحلیل و ترسیم شبکه های داده، از جمله شبکه های معنایی در متون).
۶. R (Stylometry with R package)، آدرس وبسایت: <https://www.r-project.org>، (زبان برنامه نویسی آماری متن باز که به کمک بسته هایی مانند Stylo، برای تحلیل سبک شناسی آماری و تحلیل های ادبی کاربرد دارد).

## فهرست منابع

- فیروزی مقدم، آر.ش. (۱۴۰۳). «بررسی و شناخت کارکردهای هوش مصنوعی در پردازش و تحلیل متون نثر فارسی بر پایه نثرپژوهی رایانشی». چهاردهمین همایش بین‌المللی بزرگداشت ابوالفضل بیهقی و مطالعات میان‌رشته‌ای نثر فارسی، ایران.
- مولوی، جلال‌الدین محمد (۱۳۷۸). کلیات شمس یا دیوان کبیر. با تصحیحات و حواشی بدیع‌الزمان فروزانفر. تهران: امیرکبیر.
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.
- Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. University of Chicago Press.
- Piper, A. (2018). *Enumerations: Data and literary study*. University of Chicago Press.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Burrows, J. F. (2002). 'Delta': A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 267-287.
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60.)
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.