

کاربرد علم داده در آموزش مهندسی شیمی

میترا لواسانی^۱، هومن ضیائی حلیمه جانی^۲، رحمت ستوده قره باغ^۳،

رضا ضرغامی^۴ و نوید مستوفی^۵

تاریخ دریافت: ۱۳۹۹/۴/۲۴، تاریخ پذیرش: ۱۴۰۰/۱/۸

DOI: 10.22047/IJEE.2021.239468.1759

چکیده: امروزه فناوری‌های جدید، مجموعه‌های متنوعی از داده‌ها را در علوم مهندسی تولید می‌کنند. علم داده، ابزارهای لازم برای ذخیره‌سازی، تجزیه و تحلیل و مدیریت این مجموعه داده‌ها را ارائه کرده و سبب ایجاد تحولی جدی در توسعه علم و فناوری گردیده است. در آینده‌ای نزدیک تقریباً در همه حوزه‌ها، مشاغل سنتی به سرعت با مشاغل داده‌محور جایگزین می‌شوند. نبودن نیروی انسانی کارآمد و آموزش دیده مانعی جدی برای مدیریت داده‌های حجیم در سازمان‌ها، بنگاه‌ها و دستگاه‌های اجرایی است. در نتیجه، آموزش و به‌کارگیری علم داده در مقطع کارشناسی و تحصیلات تکمیلی در رشته‌های مهندسی و از جمله مهندسی شیمی اجتناب‌ناپذیر است. در این مقاله، به بررسی اهمیت ایجاد دروس مرتبط با علم داده، آموزش روش‌های ارزیابی داده‌های کلان و نقش این آموزش‌ها در توسعه مهندسی شیمی در دانشگاه و صنعت پرداخته می‌شود. برای این منظور، دروس علم داده در دانشگاه‌های برتر مورد بررسی قرار می‌گیرد و در نهایت پیشنهادهایی برای ایجاد دروس علوم داده در برنامه آموزشی مهندسی شیمی در ایران ارائه می‌شود. هم‌چنین سرفصل‌هایی به‌منظور نحوه ارائه این درس یا دروس پیشنهاد می‌گردد. ارائه این آموزش‌ها به مهندسی شیمی، می‌تواند آن‌ها را به‌منظور کسب فرصت‌های شغلی و نحوه مواجهه با داده‌های حجیم و کلان در صنایع شیمیایی باری نماید.

واژگان کلیدی: داده‌های کلان، علم داده، آموزش مهندسی شیمی، آموزش مهندسی، مدیریت داده

۱- پژوهشگر پسا دکتری، دانشکده فنی دانشگاه تهران، ایران. mit.lavasany@yahoo.com

۲- دانشجوی دکتری، دانشکده فنی دانشگاه تهران، ایران. hooman.ziaei@ut.ac.ir

۳- استاد، دانشکده فنی دانشگاه تهران، ایران. (نویسنده مسئول) sotudeh@ut.ac.ir

۴- استاد، دانشکده فنی دانشگاه تهران، ایران. rzarghami@ut.ac.ir

۵- استاد، دانشکده فنی دانشگاه تهران، ایران. mostoufi@ut.ac.ir

۱. مقدمه

مهندسی شیمی رشته‌ای است که با حوزه‌های متعدد صنعتی و فناوری در ارتباط است. به طور کلی، مهندسان شیمی با طراحی فرایندها، مواد مورد نیاز در زندگی بشری را تولید می‌کنند. هر فرایند با آزمایش در آزمایشگاه‌ها آغاز و به دنبال آن، با کسب دانش فنی مورد نیاز، تولید و در مقیاس‌های بزرگ‌تر انجام می‌شود. مأموریت برنامه کارشناسی رشته مهندسی شیمی برای دانشجویان، افزایش دانش‌های پایه‌ای همچون فیزیک، شیمی و ریاضیات است که بنیان و اساس طراحی فرایندها در صنایع شیمیایی می‌باشند. علاوه بر این دانشجویان این رشته لازم است که همواره توانایی استفاده از روش‌ها، مهارت‌ها و ابزارهای مهندسی مدرن را فراگیرند.

در دهه گذشته، مهندسان شیمی با داده‌های پیچیده‌تری نسبت به قبل سروکار داشته‌اند. به دلیل پیشرفت و گسترش فناوری‌ها و رویکردهای جدید و همچنین دسترسی بیشتر به حسگرهای ارزان قیمت و ابزار دقیق، مجموعه‌های متنوعی از داده‌ها با سرعت بی‌سابقه‌ای در واحدهای صنعتی مدرن تولید می‌گردند (Asadi Samadi et al., 2018). علاوه بر این دانشجویان و پژوهشگران دانشگاهی با انجام آزمایش‌هایی بر مبنای رویکرد موازی سازی داده‌ها، مانند ردیابی ذرات با وضوح بالا و یا با مدل سازی ریاضی و شبیه سازی رایانه‌ای، داده‌هایی با حجم زیاد ارائه می‌دهند (Venkatasubramanian, 2009). از آنجا که آموزش مهندسی شیمی با این حجم از انفجار داده‌ها همگام نبوده است، مهندسان شیمی بیشتر داده‌ها را با استفاده از روابط ساده در داده‌ها، ایجاد نمودارهای رایج برای نمایش همبستگی‌های دومتغیره و چندمتغیره و استفاده از رویکردهای فرضیه محور، مورد تحلیل قرار می‌دادند. این روش‌ها متعلق به دوره‌ای با حجم داده کم است که در آن مهندسان می‌توانستند داده‌ها را به صورت ذهنی یا چشمی ارزیابی کنند. ولی بدون داشتن آموزش در مورد مدیریت و استخراج دانش از داده‌ها، اکثر اطلاعات مربوط به پالایشگاه‌ها و مجموعه‌های عظیم صنعتی، آزمایش‌ها و شبیه سازی‌های رایانه‌ای بی‌اثر می‌گردد (Beck et al., 2016).

به منظور بهره‌برداری مؤثر از داده‌ها با حجم روزافزون، لازم است مهندسان شیمی روش‌های علم داده را فراگیرند. هدف حوزه نوظهور علم داده، فراهم آوردن ابزاری برای مدیریت مجموعه‌های داده‌های عظیم در دسترس مهندسان و استخراج دانش از این اطلاعات انبوه با استفاده از روش‌های قدرتمند آماری است. این علم جدید به کمک پیشرفت سریع فناوری در محاسبات، می‌تواند انقلابی در طراحی، کنترل و ارزیابی فرایندهای شیمیایی ایجاد نماید. به طور کلی با پیشرفت و گسترش تحولات در زمینه علم داده، تجارت، علم و جامعه متحول می‌گردند (Marr, 2017).

در دنیای امروز آموزش علم داده اجتناب‌ناپذیر است. لذا آشنایی اعضای هیئت علمی، مدیران

و دانشجویان با علم داده برای رویارویی با این دوره مدرن داده‌محور ضروری است. این مقاله به معرفی مفاهیم داده‌های کلان و علم داده، اهمیت و کاربرد علم داده در مهندسی شیمی می‌پردازد. رویکردهای مختلفی برای تدریس علم داده در دانشگاه‌های برتر دنیا مورد بررسی قرار گرفته است و در نهایت پیشنهادهایی در مورد چگونگی گنجاندن اصول علم داده در برنامه آموزشی مهندسی شیمی در ایران ارائه می‌شود.

۱-۱. تفاوت داده و داده کلان

داده، واحد جداگانه‌ای از اطلاعات است. هر داده یک کیفیت یا کمیت از برخی از اجزا یا پدیده‌ها را توصیف می‌کند. در مدل‌های تحلیلی، داده‌ها توسط متغیرها نشان داده می‌شوند. داده کلان^۱ عبارت از داده با اندازه بزرگ است. کلان داده یا مه داده‌ها^۲ اصطلاحی است که برای توصیف مجموعه‌ای از داده‌های متنوع با حجم و سرعت بالا به کار می‌رود. به طور خلاصه، اندازه چنین داده‌هایی فراتر از حدی است که با نرم‌افزارها و روش‌های معمول بتوان آن‌ها را در یک زمان قابل قبول، دریافت، ذخیره، مدیریت و پردازش کرد. به طور کلی، تفاوت‌های اصلی داده‌های کلان و داده‌های سنتی^۳ در جدول ۱ بیان شده‌اند (Hu et al., 2014). بانک اطلاعاتی سنتی برای انتقال و مدیریت تعداد زیادی از داده‌ها، به سخت‌افزار و نرم‌افزاری پیچیده و گران قیمت نیاز دارد. در مه داده‌ها، پخش حجم انبوهی از داده‌ها بین سامانه‌های مختلف (پردازش توزیع شده و موازی)، باعث کاهش مقدار داده می‌شود. در این مسیر، سخت‌افزارهای مناسب و نرم‌افزار منبع باز، پردازش داده‌های بزرگ را بسیار آسان کرده است (Sun et al., 2014). امروزه هم‌زمان با رشد فناوری و افزایش حجم داده‌ها، هزینه‌های ذخیره‌سازی آن‌ها نیز کاهش یافته است. شکل ۱ روند تغییرات هزینه ذخیره‌سازی داده بر حسب سال را نشان می‌دهد (Years trend, 2020). همان‌طور که در این شکل نشان داده شده است، هزینه‌های ذخیره‌سازی داده‌ها با زمان، روندی نزولی دارند. علاوه بر این کاربرد فناوری رایانش ابری در داده‌های کلان سبب کاهش هزینه‌های سخت‌افزاری، پردازش و صحت‌گذاری ارزش این داده‌ها می‌گردد (Akraghilipour et al., 2016). هم‌زمان با کاهش هزینه ذخیره اطلاعات، قدرت پردازش رایانه‌ها نیز افزایش یافته است. شکل ۱ هم‌چنین ارتقای توان محاسباتی در طول ۴۲ سال را نشان می‌دهد. پردازنده‌ها در طول تاریخ خود دست‌خوش تحول عظیمی شده‌اند. این روند تأثیر قابل توجهی در رشد فناوری داده‌های کلان و ارزیابی در زمان واقعی دارد. پردازنده‌های سریع‌تر با تعداد هسته بیشتر، امکان توسعه الگوریتم‌های موازی و کاهش زمان محاسبات برخط را افزایش می‌دهند (Years trend, 2020).

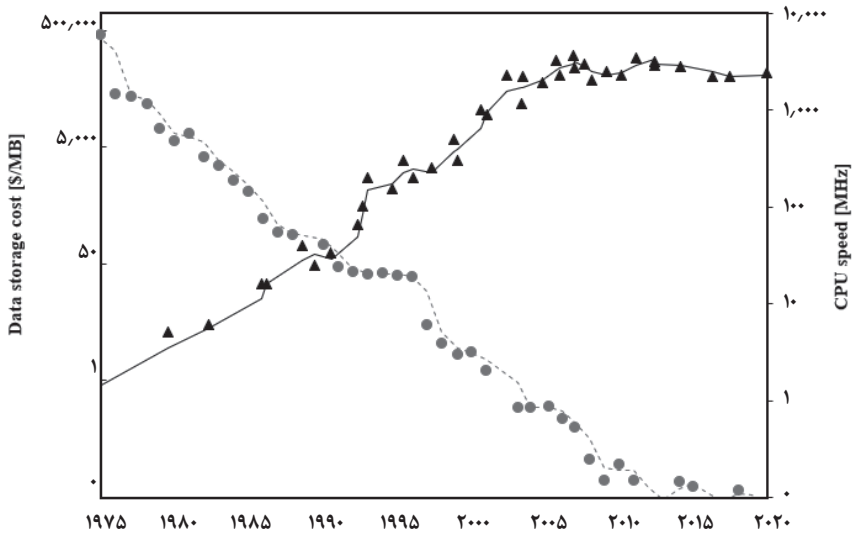
1- Big Data

۲- مه داده‌ها برابرنهاده فرهنگستان زبان و ادب فارسی است.

3- Traditional

جدول ۱. تفاوت‌های اصلی داده‌های کلان و داده‌های معمولی (Hu et al., 2014)

داده‌های کلان	داده‌های سنتی	معیار
به‌روزرسانی مداوم (در حال حاضر ترابایت ^۱ یا پتابایت ^۲)	گیگا بایت ^۳	حجم
بسیار سریع	در ساعت، در روز	نرخ تولید
ساختار نیافته یا نیمه‌ساختاریافته	ساختاریافته	ساختار
کاملاً توزیع شده	متمرکز	منبع داده
سخت	آسان	یکپارچه‌سازی داده ^۴
کمتر	بیشتر	دسترسی



شکل ۱. هزینه ذخیره‌سازی داده و سرعت محاسباتی بین سال‌های (۱۹۷۵-۲۰۲۰)

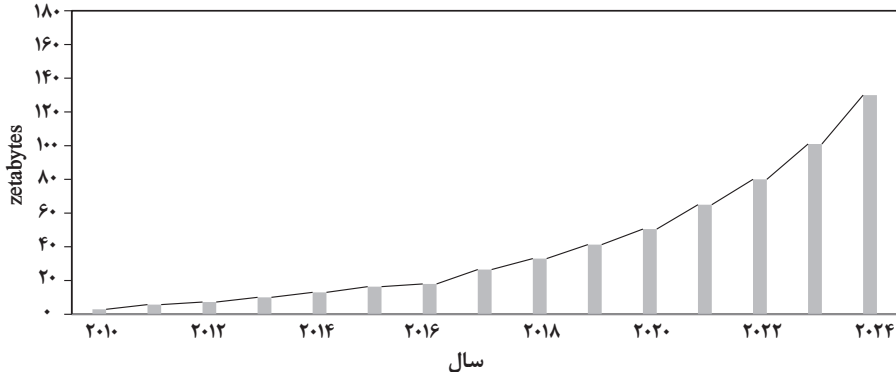
داده‌های کلان را می‌توان بر اساس خصوصیات آن‌ها تعریف کرد. هر یک از این خصوصیات یکی از ابعاد محیط کلان داده (حجم^۵، سرعت^۶ و تنوع^۷) را معرفی می‌کند. حجم، در واقع مقدار داده‌هایی است که با سرعت فزاینده‌ای روزبه‌روز در حال گسترش است (Ardeshiri Lajimi et al., 2016). در تعریف کلان داده، حجم داده مشخص نشده است چون میزان بزرگ بودن داده، هم‌زمان با پیشرفت فناوری‌های ذخیره‌سازی و پردازش اطلاعات روزبه‌روز در حال افزایش است. دنیای دیجیتال دائماً اطلاعات بیشتری را توسط تجهیزات و ابزارهای مختلف تولید می‌کند. در یک گزارش تخمین زده شده

1 - Terabyte
4 - Data Integration
7 - Variety

2 - Petabyte
5 - Volume

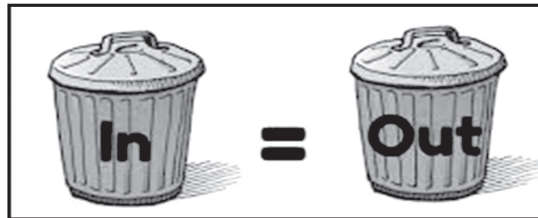
3 - Gigabyte
6 - Velocity

است که کل داده‌های دیجیتال در سال ۲۰۲۵، به ۱۷۵ زتابایت^۱ (یک زتا برابر است با یک تریلیون گیگ) خواهد رسید (شکل ۲) (Reinsel, 2017).



شکل ۲. رشد داده‌های کلان (Reinsel, 2017).

از دیگر ویژگی‌های داده‌های کلان سرعت است که بیانگر نرخ رشد حجم عظیمی از داده‌ها، جمع‌آوری و تحلیل آن‌هاست. در بسیاری از کاربردها، سرعت تولید داده‌ها از حجم آن‌ها مهم‌تر است. هر ثانیه داده‌ها افزایش پیدا می‌کنند که نه تنها باید تحلیل شوند، بلکه سرعت انتقال و دسترسی به آن‌ها نیز باید با همان سرعت انجام شود. فناوری علم داده این امکان را فراهم می‌سازد که در حین تولید، داده‌ها را بدون وارد کردن آن‌ها به پایگاه داده‌ها تجزیه و تحلیل نمود. تنوع داده‌ها به انواع داده‌های موجود از جمله داده‌های ساختاریافته، ساختارنیافته و نیمه‌ساختاریافته اشاره دارد. هم‌چنین داده‌های کلان از منابع مختلف با فرمت‌های متفاوت گردآوری می‌شوند و داده‌ها باید به گونه‌ای سازماندهی گردند تا معنا پیدا کنند. ویژگی چهارم صحت^۲ است که شامل دسترسی پذیری و پاسخگو بودن است. با توجه به این که داده‌ها از منابع مختلف دریافت می‌شوند، احتمال وجود داده‌های نادرست در آن‌ها نیز افزایش می‌یابد. در نتیجه اگر ورودی قابل اعتماد نباشد، نمی‌توان به اطلاعات استخراج شده از آن اعتماد کرد (شکل ۳) (Duever, 2019).



شکل ۳. تأثیرگذاری داده‌های ورودی بر نتایج خروجی

۲. مفاهیم علم داده

علم داده، دانشی متشکل از نظریه‌ها و مباحث مرتبط با چند رشته علمی مختلف است که از روش‌های علمی، فرایندها، الگوریتم‌ها و سامانه‌های علمی برای کسب بینش از داده‌ها در اشکال گوناگون (ساختاریافته و ساختارنیافته) استفاده می‌کند (شکل ۴) (Cao, 2018). دانشمند داده باید در زمینه‌های مختلفی مهارت داشته باشد که مهم‌ترین آن‌ها مفاهیم آماری، علوم رایانه‌ای، مهندسی داده، یادگیری ماشین، تحلیل‌های پیش‌بینی و تحلیل‌های کسب‌وکار است. در علم داده سعی بر آن است تا با واکاوی و تجزیه و تحلیل داده‌ها، اطلاعات ارزشمندی استخراج شود (Piccione, 2019). اما مهم‌تر از آن اطلاعاتی است که دانشمند داده با استفاده از انواع مدل‌های توصیفی، پیش‌بینی و استنباط می‌کند و براساس عوامل مختلف، ریشه‌یابی و پیش‌بینی مشکل‌ها، راه‌حل مناسب برای فعالیت‌های آتی سازمان‌ها و صنایع را ارائه می‌نماید (Kelleher & Tierney, 2018).



شکل ۴. علم داده، دانشی میان‌رشته‌ای

روند صعودی علاقه به علم داده توسط بسیاری از عوامل، از جمله موارد زیر به وجود آمده است (Earnshaw et al., 2019):

۱. افزایش فرایند دیجیتالی شدن برای جمع‌آوری داده‌های حجیم
۲. تجزیه و تحلیل داده‌های حجیم و پیچیده برای ایجاد فرصت‌های تجاری
۳. افزایش توان محاسباتی و ذخیره اطلاعات به کمک سامانه‌های محلی و ابری
۴. اتصال ابزارهای هوشمند به اینترنت برای جمع‌آوری داده‌های حجیم

علم داده معمولاً به سه دسته تقسیم می‌شود: مدیریت داده‌ها، روش‌های آماری یادگیری ماشین و مصورسازی. در هر دسته، روش‌ها و ابزارهای بسیاری وجود دارند که می‌توانند توسط مهندسان شیمی برای رفع چالش‌های مربوط به داده‌ها مورد استفاده قرار گیرند (Ayers, 2017). مدیریت داده‌ها به ابزارها و روش‌های سازماندهی، مرتب‌سازی و پردازش مجموعه داده‌های بزرگ و پیچیده و امکان ارزیابی در زمان واقعی^۱ جریان داده‌ها از حسگرها، ابزارها و شبیه‌سازی‌ها اشاره دارد. مدیریت داده پایه و اساس علم داده را تشکیل می‌دهد. نحوه سازماندهی، ذخیره و پردازش داده‌ها به طور قابل توجهی بر کیفیت تحلیل‌ها، سهولت به اشتراک‌گذاری و مصورسازی نتایج تأثیر می‌گذارد.

علم داده شامل حوزه‌های مرتبط با یادگیری ماشین و داده‌کاوی است. یادگیری ماشین ممکن است به عنوان یک زمینه از هوش مصنوعی تلقی شود و یک برنامه رایانه‌ای را قادر سازد تا بر اساس استنباط یا الگوهای موجود در مجموعه داده‌ها، نتیجه‌گیری کند. یادگیری ماشین به طور نزدیکی با آمار محاسباتی مرتبط است و اغلب با آن هم‌پوشانی دارد. تمرکز این شاخه نیز پیش‌بینی توسط رایانه است و پیوند محکمی با بهینه‌سازی ریاضی دارد. روش‌های یادگیری ماشین به دو دسته یادگیری نظارت‌شده^۲ و یادگیری بدون نظارت^۳ تقسیم می‌شوند. در یادگیری نظارت‌شده، هدف ایجاد یک مدل پیش‌بینی شده است که خروجی‌ها بر اساس ورودی پیش‌بینی می‌شوند. در یادگیری بدون نظارت، هدف تعیین ساختار اساسی داده‌ها بر اساس برخی ویژگی‌ها است. داده‌کاوی^۴ از روش‌های یادگیری ماشین و آمار برای کشف الگوهای مهم و جالب توجه در میان حجم انبوه داده‌ها استفاده می‌کند. منافع تجاری بسیاری در داده‌کاوی است که از جمله می‌توان به تحلیل بازار^۵ و کنترل تولید^۶ اشاره کرد (Earnshaw et al., 2019).

آخرین و مهم‌ترین مؤلفه علم داده، مصورسازی^۷ داده‌ها است. روش‌های مصورسازی، روش‌های مورد استفاده برای ایجاد جداول، تصاویر، نمودارها و دیگر روش‌های نمایش بصری برای درک داده‌ها هستند. قابل ذکر است که مصورسازی، جایگزین الگوریتم‌های آماری نیست بلکه از این روش‌ها برای کمک به تصمیم‌گیری استفاده می‌گردد. در حقیقت، هدف از نمایش گرافیکی، ارائه درک عمیق‌تر از نتایج ارزیابی داده‌ها و حتی کشف برخی از همبستگی‌های پنهان است. بینایی قدرتمندترین ابزار ارتباطی انسان است و درک ویژگی‌ها و ساختارها از نمودارها سریع‌تر از متن است. با این حال ظرفیت بینایی انسان به دو یا سه بعد محدود است. هنگامی که تجسم داده‌ها در مقیاس بزرگ مورد نظر است، باید با استفاده از روش‌های کاهش ابعاد به طور قابل توجهی اندازه داده‌ها را قبل از ارائه داده‌های واقعی کاهش داد (Tiwari, 2018).

1- Real-time processing

2- Supervised learning

3- Unsupervised learning

4- Data mining

5- Market analysis

6- Production control

7- Visualization

۳. علم داده و مهندسی شیمی

مهندسان شیمی به دلیل گسترش داده‌های حاصل از حسگرها و ابزار دقیق، قابلیت اندازه‌گیری جدید با توسعه اینترنت اشیاء و حسگرهای هوشمند و بهبود توان ذخیره‌سازی داده‌ها، مانند رایانش ابری، با داده‌های عظیم و پیچیده‌تری نسبت به گذشته سروکار دارند. به منظور بهره‌برداری از داده‌های موجود، مهندسان شیمی باید از روش‌های علم داده استفاده کنند (Donoho, 2017).

امروزه دانشجویان یا محققان در زمینه تحقیق و اکتشاف، آزمایش‌هایی را با حجم بالایی از داده، مانند ردیابی ذرات با وضوح بالا، انجام می‌دهند که ممکن است در یک روز اطلاعات بیشتری را نسبت به یک دوره دکتری در دهه‌های قبل تولید کنند. هم‌چنین مهندسانی که مدل‌سازی ریاضی و شبیه‌سازی رایانه‌ای انجام می‌دهند، به دلیل توسعه الگوریتم‌های پیشرفته و سخت‌افزارها، اکنون شبیه‌سازی‌هایی با وضوح بی‌سابقه انجام و مجموعه داده‌های عظیمی را ارائه می‌دهند (White, 2016). نمونه‌های زیادی از کاربرد روش‌های علم داده در جامعه مهندسی شیمی وجود دارند که سال‌هاست استفاده می‌گردند، همانند ارزیابی چندمتغیره، تشخیص عیب برخط^۱، حسگرهای استنباطی^۲، ارزیابی داده‌های دسته‌ای^۳، طراحی آزمایش‌ها و برازش (Ayers, 2017). مهندسان هم‌چنین از ابزارهای علم داده برای نظارت و کنترل فرایندها، پیش‌بینی ساختارها و خواص شیمیایی و اطلاع‌رسانی در مورد تصمیم‌های تجاری و تحقیقاتی استفاده می‌کنند (Chiang et al., 2017).

صنایع شیمیایی همواره برای حفظ شرایط بهینه و ایمن از کنترل فرایندها استفاده می‌کنند. کنترل خودکار فرایند و حفظ دقیق محدودیت‌های متغیرها برای فرایندهای تولید پیوسته و ناپیوسته ضروری است. هم‌چنین فرایندهای صنعتی از سامانه‌های نظارتی برای اطمینان از کیفیت و ثبات بهره‌مند می‌شوند. امروزه ظهور فناوری‌های جدید و ارزیابی داده‌های کلان منجر به توسعه سامانه‌های کنترل و نظارت فرایندها شده است. در فرایندهای مدرن، کارکرد ایمن و پایدار و کنترل حوادث غیرطبیعی اهداف اصلی سامانه‌های پیچیده تولید است و فناوری‌های مبتنی بر داده، در حال تغییر اتوماسیون فرایند هستند (Chiang & Colegrove, 2007; Abonyi, 2013). تشخیص زودهنگام عیب به تصمیم‌گیری در شرایط واقعی کمک می‌کند و به سرعت ناهنجاری‌های موجود را تشخیص می‌دهد. فرایندهای شیمیایی رویکرد مبتنی بر داده‌ها (رویکرد دانش‌محور) را برای استخراج دقیق ویژگی‌های عیب تقاضا می‌کنند. تا کنون از الگوریتم‌های مختلفی برای پیش‌بینی حوادث غیرطبیعی و اتفاق‌ها استفاده شده است، با این حال داده‌کاوی به‌طور گسترده برای شناسایی الگوها و ساختارها در صنایع شیمیایی استفاده می‌گردد (Venkatasubramanian, 2003). از دیگر مزایای ارزیابی داده‌های کلان، حسگرهای

مجازی^۱ هستند که برای جایگزینی اندازه‌گیری واقعی متغیرها با اندازه‌گیری مجازی با روش‌های داده‌محور مدل می‌شوند. حسگر مجازی کیفیت محصول یا متغیرهای بحرانی اندازه‌گیری نشده را بر اساس سایر متغیرهای اندازه‌گیری شده با فناوری ترکیبی (مثلاً مدل ریاضی، پردازش داده‌ها و روش‌های نرم‌افزاری) تخمین می‌زند (Kano & Fujiwara, 2012). این فناوری نوظهور، به دلیل قابلیت تخمین استنباطی، کاربردهای زیادی در صنعت دارد (Abonyi, 2014).

با رشد فناوری و پیچیده شدن فرایندهای شیمیایی، صنایع برای ساماندهی و ذخیره داده‌ها به روش‌های مدل‌سازی داده با توان بالا نیاز دارند. بنابراین جای تردید نیست که مدل‌ها و شبیه‌سازهای قبلی، قادر به مدل‌سازی داده‌های کلان نیستند و محققان برای غلبه بر مشکل‌های همگرایی مدل‌ها، از الگوریتم‌های تخصصی استفاده می‌کنند. به عنوان مثال می‌توان به کنترل و بهینه‌سازی سامانه گازی سازی زغال سنگ^۲ (Guan, 2017) و هم‌چنین مدل‌سازی انتشار گازهای گلخانه‌ای (NOx) از یک بستر سیال اشاره کرد (Ma et al., 2018). دینامیک سیالات محاسباتی^۳ روشی معمول برای طراحی و ارزیابی فرایندهای مهندسی است، گرچه برای جریان‌های چندفاز بسیار گران و زمان‌بر است. اخیراً از روش‌های متداول در زمینه هوش مصنوعی^۴ و داده‌کاوی برای ساختن یک مدل هوشمند CFD استفاده شده است. مدل داده‌محور برای بررسی جریان چندفازی و رفتار ذرات جامد و گاز در یک بستر سیال مورد استفاده قرار گرفته است. این روش می‌تواند نتایج شبیه‌سازی CFD را با دقت معقول و زمان محاسبه کمتر تولید کند (Ansari, 2019). علاوه بر این در فرایندهای پیچیده، با افزایش تعداد مولفه‌های تصمیم‌گیری و وجود نوفه^۵ها، شبیه‌سازهای فعلی و روش‌های موجود در بهینه‌سازی فرایندها مؤثر نیستند و لازم است از روش‌های یادگیری ماشین استفاده گردد (Cozad et al., 2014).

مهندسان شیمی از طریق تفسیر داده‌های بزرگ توانسته‌اند فرایندهای جدیدی را اختراع کنند، سود و بهره‌وری عملیاتی را افزایش دهند و با سایر صنایع همکاری کنند. هم‌چنین افزایش داده منجر به درک بهتر مسائل، توانمندسازی و تسهیل تصمیم‌گیری شده است (Ayers 2017). ادغام زودهنگام داده‌های بزرگ در مهندسی شیمی مزایای بسیاری داشته و به تعدادی از سازمان‌ها در کسب مزایای رقابتی کمک کرده است (Chiang et al., 2017). با این حال، کاربردهای علم داده در مهندسی شیمی نسبتاً محدود بوده و بسیاری از فرصت‌ها برای پیشرفت این رشته ناشناخته مانده است. کاربرد ارزیابی داده‌های بزرگ در مهندسی شیمی در حال رشد است. شکل ۵ تعداد مقاله‌های

1 - Soft sensors

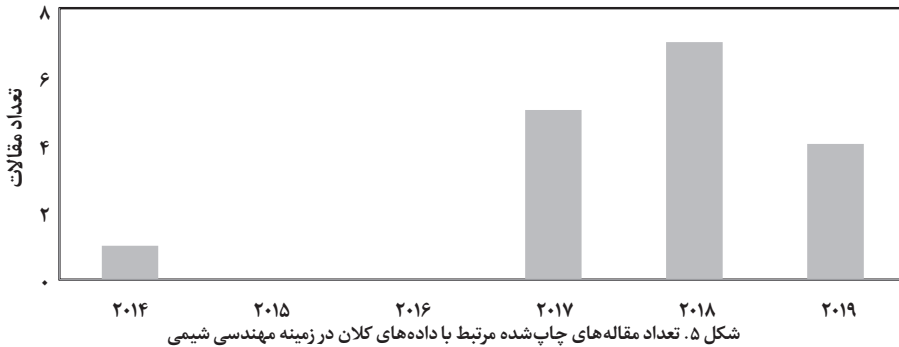
2 - Coal gasification system

3 - Computational fluid dynamics

4 - Artificial intelligence

۵- نوفه برابر نهاده فرهنگستان برای نویز است.

چاپ شده در این زمینه از سال ۲۰۰۰ را نشان می دهد. با توجه به این شکل، که بیانگر افزایش علاقه به داده های کلان در سراسر دنیاست، لازم است این فناوری به عنوان یک مفهوم جدید به دانشجویان، که مهندسان آینده این کشور هستند، آموزش داده شود.



۱-۳. ضرورت ایجاد درس علم داده در آموزش مهندسی شیمی

مهندسان شیمی زمینه بسیار قوی در ریاضیات و حل مسئله دارند. لذا به خوبی آمادگی فراگیری علم داده را دارند. آن ها به تعداد کمی از پتانسیل هایی که داده های بزرگ ارائه می دهند نیاز دارند، بنابراین از شاخه های خاصی استفاده می کنند. مهندسان شیمی باید نحوه ارزیابی داده های کلان را فراگیرند، در غیر این صورت ممکن است به نتایج نادرستی دست یابند. دروس علم داده دانشجویان را با ابزارهای علم داده و روش های مدل سازی محاسباتی مرتبط با مهندسی شیمی مدرن آشنا می سازد. مهارت های مورد نیازی که باید در آموزش مهندسی شیمی گنجانده شوند، عبارتند از: یادگیری ماشین بر اساس آمار، از جمله آمار استنباط بیزی^۱ و ارزیابی چندمتغیره^۲، علوم رایانه شامل ساختار داده ها، الگوریتم ها و سامانه ها شامل محاسبات توزیع شده^۳، بانک های اطلاعاتی، محاسبات موازی و محاسبات متحمل به عیب^۴، آگاهی و دانش از همبستگی^۵ و علیت^۶ و نهایتاً توانایی حل مشکل ها به شکلی که منجر به راه حل های مؤثری گردند (Dhar, 2013).

بسیاری از روش های به کار رفته در علم داده، برای مهندسان شیمی آشنا هستند. با این حال، به طور کلی تمامی روش های مورد نیاز برای انجام پروژه های علم داده، در برنامه درسی کارشناسی مهندسی شیمی پوشش داده نشده اند. از دیدگاه سازمان ها، عدم وجود دانش و فقدان نیروی کار

1 - Bayesian statistics

2 - Multivariate analysis

3 - Distributed computing

4 - Fault-tolerant computing

5 - Correlation

6 - Causation

آموزش دیده، از جمله موانع اجرای موفقیت آمیز علم داده است. ذی نفعان در دانشگاه، دولت و صنعت به منظور بهره‌وری از مزایای این فناوری جدید، سعی در آموزش آن به مهندسان شیمی دارند که در واقع نیروی کار آینده صنایع را تشکیل می‌دهند. فاصله صنعت ایران از فناوری‌های مدرن دنیا و عد تدریس واحدهای کاربردی مرتبط با صنعت موجود در کشور از یک طرف و گستردگی صنایع مرتبط با رشته مهندسی شیمی از طرف دیگر، باعث افزایش شکاف میان آموزش‌های دانشگاهی و مهارت‌های لازم در صنعت شده است. دانشجویان مباحثی از علوم نظری، طراحی و محاسبات تجهیزات پیشرفته را آموزش می‌بینند که عملاً در صنایع داخلی مورد استفاده قرار نمی‌گیرند و از طرفی، بسیاری از مفاهیم ابتدایی در دانشگاه تدریس نمی‌شوند و دانش‌آموختگان مهندسی شیمی فاقد مهارت‌های کاربردی مورد نیاز شرکت‌های مهندسی هستند (Kheradmandi Nia et al., 2019).

امروزه حجم عظیمی از داده در صنایع شیمیایی جمع‌آوری می‌شود. از آن جایی که کشف اطلاعات موجود در داده‌ها حیاتی است، شرکت‌ها علاقه‌مند به استخدام مهندسان شیمی آشنا با مفاهیم علم داده هستند. به عنوان مثال، امروزه در مجتمع‌های صنعتی مدرن یا سامانه‌های مهندسی از قبیل موتورهای هوا و نقلیه، سامانه‌های تولیدی، شبکه‌های قدرت، ماشین‌های الکتریکی، سامانه‌های تبدیل انرژی و هم‌چنین مجتمع‌های فرایندهای شیمیایی و زیست‌شناختی، ایمنی فرایند^۱ و کیفیت محصول‌ها از مهم‌ترین چالش‌های موجود است که علاقه‌مندی قابل توجه‌ای را، چه در حوزه تحقیقاتی و چه در حوزه صنعتی، برانگیخته است. در این میان پایش و کنترل فرایند^۲ از مهم‌ترین ابزارهای مورد استفاده به منظور رسیدن به ایمنی مورد قبول، کیفیت محصول‌های بالا، کاهش هزینه‌ها و هم‌چنین عملکرد مورد اطمینان فرایند در طولانی مدت است. در واقع پایش فرایند از این جهت مورد اهمیت است که در هر فرایندی، امکان بروز عیب^۳ یا شرایط غیرعادی اجتناب‌ناپذیر است. در نتیجه تشخیص^۴ و شناسایی^۵ دقیق پتانسیل‌های بروز شرایط غیرعادی و هم‌چنین به‌کارگیری سامانه‌ای به منظور کاهش توان تخریبی این شرایط یا حتی جلوگیری از آن ضروری است. با توجه به حجم بالای داده‌های تولید در یک مجتمع پالایشگاهی یا پتروشیمی یا هر مجتمع شیمیایی دیگر، تسلط بر علم تحلیل داده و خصوصاً داده‌های کلان، در کنار مهندسی شیمی از اهمیت به‌سزایی برخوردار است (Ziaei-Halimejani et al., 2021). استفاده از روش‌های علم داده، نیاز به مهارت‌هایی بیشتر از برنامه‌فعلی مهندسی شیمی دارد. یک گزینه برای پرداختن به این امر، تنظیم برنامه درسی با تغییر دروس موجود و معرفی دروس انتخابی است. راه حل دیگر شامل معرفی یک درس علم داده در مقطع کارشناسی است. این درس باید علاوه

1 - Process safety
4 - Detection

2 - Process monitoring and control
5 - Diagnosis

3 - Fault

بر نحوه مدیریت داده، چگونگی استخراج مفاهیم و توانایی تصمیم‌گیری درست در مورد داده‌ها را نیز آموزش دهند. لحاظ کردن این مفاهیم در برنامه درسی، امکان یادگیری ابزارهای مورد نیاز برای حل مسئله در دنیای واقعی را فراهم می‌کند (Beck, 2017).

۴. آموزش علم داده

فرصت‌های آموزشی علم داده برای دانشجویان، به سرعت در حال رشد است. دانشگاه‌ها، دانشکده‌ها، انجمن‌ها و سازمان‌های دیگر ارائه طیف وسیعی از برنامه‌ها را برای دانشجویان با علایق مختلف آغاز کرده‌اند. آموزش علم داده در مقطع کارشناسی، در حال حاضر به اشکال مختلف ارائه می‌شود و انتظار می‌رود این تغییر در آینده نزدیک نیز ادامه یابد. روش‌های معمول شامل موارد زیر است:

- ارائه مقدماتی علم داده: از طریق یک درس برای ارائه آموزش عمومی
 - کارشناسی ارشد علم داده: شامل مهارت‌های پیشرفته به عنوان یک رشته تحصیلی
 - آموزش جزئی در علم داده: آموزش مهارت‌های مورد نیاز مرتبط با هر رشته تحصیلی
 - مدارک و گواهی‌نامه‌های دوساله (کاردانی)
 - گواهی‌نامه‌های دیگر
 - دوره‌های برخط: که می‌تواند تعداد زیادی از دانش‌آموزان را در سطوح مختلف درگیر کند
 - ارائه برنامه‌های تابستانی به عنوان دوره کارآموزی
- مؤسسه‌های دانشگاهی قبل از اضافه کردن دوره‌های علم داده و برنامه‌نویسی به دوره کارشناسی، باید عوامل مختلف مانند نیازهای دانشجویان، توانمندی‌های دانشکده و استادان، شیوه‌های مناسب آموزش و منابع را بررسی نمایند و سپس متناسب با آن‌ها برنامه درسی را تغییر دهند (National Academies of Sciences, 2018).

چندین مؤسسه آموزشی دوره‌های مقدماتی علم داده برای آموزش دانشجویان را برگزار می‌کنند و احتمالاً بسیاری از مؤسسه‌های دیگر نیز این دوره‌ها را در سال‌های بعد ایجاد خواهند کرد. دوره‌های مقدماتی، دانشجویان را به علم داده علاقه‌مند می‌سازد و آن‌ها را به این رشته جذب و به طور گسترده‌ای برای رویایی با دنیایی مبتنی بر داده، آماده می‌کند. برخی از مؤسسه‌ها الزامات عمومی مورد نیاز در علم داده را با مشارکت چندین گروه تدریس می‌نمایند. این انعطاف‌پذیری تضمینی برای امکان ادغام علم داده در سایر برنامه‌های درسی است. مهارت‌های قابل انتقال علم داده شامل آمار توصیفی، تجسم، نمونه‌گیری، برنامه‌نویسی، مدیریت پایگاه داده، استنتاج و ارزیابی است. به نظر می‌رسد دانشجویان برای ثبت نام در دوره‌های علم داده مشتاق باشند زیرا با یادگیری این مفاهیم به ارزش مدارک خود اضافه می‌کنند. برخی از دانشگاه‌های ارائه‌دهنده دوره مقدماتی علم داده در جدول ۲ آمده است.

جدول ۲. دانشگاه‌های ارائه‌دهنده دوره مقدماتی علم داده

دانشگاه	توضیحات
دانشگاه کالیفرنیا، برکلی ^۱	<ul style="list-style-type: none"> - ایجاد دروس بین‌رشته‌ای در دانشکده‌های آمار و رایانه با مشارکت رشته‌های مختلف - ارائه مبانی علم داده در دانشکده‌های فوق برای همه دانشجویان بدون پیش‌نیاز - رشد تعداد دانشجویان از ۱۰۰ نفر در ۲۰۱۵ به ۱۱۰۰ نفر در ۲۰۱۸
دانشگاه کارنگی ملون ^۲ ، یک دانشگاه تحقیقاتی خصوصی در پیتسبورگ ^۳ ، پنسیلوانیا ^۴	<ul style="list-style-type: none"> - عنوان دوره "استدلال با داده" - اولین دوره در آمار و علم داده با تمرکز بر مفاهیم - استفاده از یک پلتفرم تعاملی^۵ برای ارزیابی دانشجویان - دوره‌ای شامل چندین پروژه تحلیل داده دانشجو محور - درسی عمومی مورد نیاز برای کلیه دانشجویان دانشکده‌های علوم انسانی و علوم اجتماعی
دانشگاه واشنگتن ^۶ ، یک دانشگاه تحقیقاتی عمومی در سیاتل ^۷ .	<ul style="list-style-type: none"> - ارائه مقدمه‌ای بر علم داده برای دانشکده‌های آمار، علوم رایانه، مهندسی و دانشکده اطلاعات - آموزش مفاهیم جمع‌آوری و مدیریت داده‌ها، مصورسازی، استنباط آماری و یادگیری ماشین - فراگیری مبانی علم داده از طریق کتاب تفکر محاسباتی و استنباطی
کالج مونتگومری ^۸	<ul style="list-style-type: none"> - ارائه مقدمه‌ای بر علم داده - یادگیری روش‌های جمع‌آوری، سازماندهی، مدیریت، ارزیابی، آماده‌سازی و مصورسازی داده‌ها

دوره علم داده مقدماتی می‌تواند دانشجویان را به دوره‌های تکمیلی جذب نماید. دانشجویان می‌توانند از سایر رشته‌ها، همچون علوم انسانی، علوم اجتماعی، هنر، فناوری، مهندسی و ریاضیات باشند. دوره‌های مقدماتی علم داده پیش‌نیازی ندارد و مهارت‌ها و تجارب قابل استفاده در طیف گسترده‌ای از مشاغل آینده را در اختیار دانشجویان می‌گذارند. این دوره‌ها دانشجویان را با نقش ارزشمند تخصص علم داده در نیروی کار آینده آشنا می‌سازند. دوره‌های مقدماتی با گذشت زمان تکامل و بهبود یافته و با توسعه برنامه درسی نوآورانه می‌توان به مسائل دنیای واقعی پرداخت و این مهم، نیازمند آموزش استادان و گسترش ارتباطات بین‌رشته‌ای می‌باشد.

رشته‌های علم داده در حال ظهور در مؤسسه‌های دانشگاهی هستند و این روند در سال‌های آینده نیز ادامه خواهد داشت. مشابه دوره‌های مقدماتی علم داده، در ساختار برنامه‌ها، اهداف و محتوای این رشته نیز تفاوت‌های چشمگیری وجود دارد. برخی از رشته‌های علم داده، به‌عنوان برنامه‌های مستقلی ظهور کرده‌اند، در حالی که برخی دیگر به‌عنوان تخصص در یک رشته خاص ظاهر شده‌اند. اکثر رشته‌های علم داده کنونی شامل دروس ریاضیات، آمار و علوم رایانه هستند. مباحث متداول تحت عنوان "علم داده" عبارتند از: ارزیابی داده‌های پیشرفته، مه‌داده‌ها، داده‌کاوی، مدل‌سازی و تفکر محاسباتی. بسیاری از رشته‌های علم داده شامل دروس اجباری یا انتخابی از خارج گروه اصلی، معمولاً

1 - <http://www.ds100.org/sp20/>. Accessed on May 2020.

3 - Pittsburgh

6 - University of Washington

4 - Pennsylvania

7 - Seattle

2 - Carnegie Mellon University

5 - Interactive platform

8 - Montgomery College

اقتصاد، تجارت، روان‌شناسی، زیست‌شناسی و جغرافیا هستند. برخی از رشته‌های علم داده نیز برای کمک به تقویت مهارت‌ها، به یک دوره عملی نیاز دارند (National Academies of Sciences, 2018). در حال حاضر، بسیاری از رشته‌های چهارساله علم داده در یکی از این سه دسته قرار می‌گیرند: (۱) رشته‌های مستقر در دانشکده اقتصاد (یعنی برنامه‌هایی در تجزیه و تحلیل تجارت که بیشتر شامل کلاس‌های بازاریابی و امور مالی بوده و دروس محاسباتی و ریاضیات در آن کمتر است). (۲) رشته‌هایی که در یکی از گروه‌های ریاضیات یا آمار تدریس می‌شوند (به‌عنوان مثال، ریاضیات و آمار بیشتر از حد متوسط نیاز با تعداد کمتری دوره‌های محاسباتی). (۳) برنامه‌هایی در گروه علوم رایانه، به‌عنوان یک رشته مستقل یا درسی در فناوری اطلاعات (به‌عنوان مثال، دوره‌های محاسباتی بیشتر، اما به‌طور بالقوه دوره‌های ریاضی کمتر). دروس ارائه‌شده در رشته علم داده در دانشگاه‌های مختلف با یکدیگر تفاوت بسیار دارند. رشته‌های چهارساله علم داده ترکیبی از این سه مدل هستند که به‌صورت مشترک توسط چندین گروه اداره می‌شوند. دانشگاه کالیفرنیا، از جمله دانشگاه‌هایی است که اصول و روش‌های علم داده را در سطح بالایی معرفی می‌کند (University of California, 2020). در این دوره، محاسبات و تفکر استنباطی در هم ادغام شده و دروس اصلی مقدماتی علم داده و ریاضیات آموزش داده می‌شود. این دوره پیشرفته، به‌جای تکیه بر ابزارها، فناوری‌های جدید علم داده را ارائه می‌دهد. هم‌چنین دوره‌های پیشرفته علوم رایانه و آمار مربوط به علم داده را بدون هیچ پیش‌نیازی به دانشجویان آموزش می‌دهد. این مباحث شامل زبان‌هایی برای تبدیل، بررسی و ارزیابی داده‌ها، الگوریتم‌های یادگیری ماشین از جمله برازش، طبقه‌بندی و خوشه‌بندی، اصول تصویرسازی اطلاعات، مفاهیم آماری، خطای اندازه‌گیری و روش‌های پردازش داده است.

برای آن دسته از دانشجویان مقطع کارشناسی ارشد که علاقه‌ای به گذراندن دوره چهارساله علم داده ندارند، می‌توان یک دوره کوتاه‌مدت برای آموزش مهارت‌های علم داده ارائه کرد. این رویکرد مناسبی برای دانشجویان سایر رشته‌هاست که تمایل دارند ضمن تکمیل رشته انتخابی خود، دانش علم داده را به‌دست آورند. در چنین برنامه‌هایی، مفاهیم علمی، ابزارها و فنون مرتبط با رشته انتخابی دانشجویان آموزش داده می‌شوند. این دوره‌ها دانش وسیعی از علم داده را به دانشجویان انتقال می‌دهند، به‌طوری که آن‌ها را برای نیروی کار آینده آماده‌تر می‌سازند. این دانشجویان نه تنها در یک رشته خاص تخصص دارند، بلکه مهارت و دانش ارزشمندی نیز کسب کرده‌اند که می‌توانند برای انواع مسیرهای شغلی در حال ظهور کاربرد داشته باشند. اگرچه دوره‌های کوتاه‌مدت علم داده گزینه مناسبی برای جمعیت خاصی از دانشجویان است، چنین برنامه‌هایی ممکن است از نظر زمانی و مکانی برای کمک به دانشجویان در توسعه مبانی علمی و مهارت‌های لازم محدود باشند.

بسیاری از مؤسسه‌ها با دوره‌های دوساله (کاردانی)، علم داده را با استفاده از رویکردهای متنوع آموزشی شروع کرده‌اند، از چند دوره منتهی به یک گواهی‌نامه تا مدرک دانشگاهی کامل که به دو سال

تحصیل نیاز دارد. ترکیبی از آمار، علوم رایانه و برنامه‌های کاربردی که متناسب با برنامه بسیار متغیر است. این برنامه‌ها اغلب به‌طور هم‌زمان خدماتی را ارائه می‌دهند: (۱) یک نقطه آغاز برای جذب طیف گسترده‌ای از جمعیت دانشجویی به علم داده هستند (۲) به اعضای موجود نیروی کار اجازه می‌دهد برای تکمیل آموزش و تجربه خود مهارت‌های جدیدی را به دست آورند (۳) با آموزش مهارت‌های خاص یا عمومی، دانشجویان گواهی‌نامه یا مدرک کاردانی را دریافت می‌کنند (۴) فرصت‌هایی را برای دانش آموزان دبیرستانی فراهم می‌سازد تا فراگیری علم داده را زودتر آغاز کنند. اکثر این اهداف، آموزش دوره کارشناسی را پشتیبانی می‌کنند و در عین حال نیازهای خاص صنعت را نیز هدف قرار می‌دهند. به‌عنوان مثال، یک برنامه مدیریت داده ممکن است تا حد زیادی روی سامانه‌های داده و جمع‌آوری متمرکز باشد، یک برنامه تجزیه و تحلیل داده‌ها ممکن است بیشتر تمرکز بر الگوریتم‌ها، آمار و یادگیری ماشین داشته باشد و یک برنامه ارزیابی تجاری ممکن است روی موضوع‌های مربوط به مشاغل و زنجیره تأمین تمرکز کند. در حال حاضر، بیشتر برنامه‌ها دوره‌هایی را برای مصورسازی داده‌ها ارائه می‌دهند اما برنامه‌های نسبتاً معدودی دوره‌هایی را برای نوشتن اطلاعات (ارتباط در علم داده) برگزار می‌کنند. برخی از این برنامه‌ها فقط به صورت برخط هستند. برخی از دوره‌ها بر آموزش مهارت‌های فنی مورد نیاز صنعت و تحقیقات، از طریق ارائه گواهی‌نامه‌های علم داده تمرکز دارند. برخی دیگر از دوره‌های دوساله مدرک کاردانی در علم داده ارائه می‌دهند که دانشجویان را برای برنامه‌های چهارساله آماده می‌کنند (CCAC, Nashua, 2020). مؤسسه‌هایی در سراسر ایالات متحده برنامه‌هایی را در زمینه علم داده و با ارائه گواهی‌های دوساله (کاردانی) ایجاد کرده‌اند. به‌عنوان مثال، دانشگاه مونتگومری مریلند^۱ به تازگی یک دوره دوساله (کاردانی) برگزار می‌کند و دروس ریاضیات و آمار، اصول رمزگذاری، نوشتن و ارتباط در علم داده و یک پروژه صنعتی با روش‌های داده‌کاوی را ارائه می‌دهد (Montgomery Certificate 256, 2020).

سایر گواهی‌نامه‌های علم داده در مؤسسه‌های دانشگاهی و صنعت، از جمله در مقطع کارشناسی، در حال گسترش هستند. هم‌چنین، دوره‌های برخط نیز رایج شده‌اند و جایگزینی برای یادگیری حضوری هستند که در این میان می‌توان به مجموعه آموزش برخط Coursera اشاره کرد. در حالی که دوره‌های صدور گواهی‌نامه می‌تواند یک منبع ارزشمند در آموزش نیروی کار موجود باشند اما روش‌های آموزشی و برنامه‌های ارائه شده، بسیار متنوع بوده و مقایسه آن‌ها اغلب دشوار است. فقدان استانداردهای دوره‌ها و آموزش‌ها می‌تواند نحوه ارزیابی کارفرمایان و دانشجویان آینده را با مشکل مواجه سازد.

۴-۱. درس دارای مفاهیم علم داده در برنامه درسی رشته مهندسی شیمی در چند دانشگاه معتبر دنیا توسعه روزافزون علوم تخصصی مرتبط با مهندسی شیمی و ایجاد توازن میان واحدهای درسی و پیشرفت‌های صنعتی از دغدغه‌های اصلی دانشکده‌های بزرگ مهندسی شیمی در دنیا است. همانند بسیاری از رشته‌های دیگر، مهندسی شیمی به‌منظور به‌کارگیری بیشتر داده‌ها، باید فناوری جدید داده‌های کلان و علم داده را استفاده کند و دانش را از داده‌ها استخراج کند. اهمیت و حساسیت این رشته در دنیا باعث شده است که دانشکده‌های مهندسی شیمی در دانشگاه‌های معتبر، به تربیت نیروی متخصص علم داده بپردازند. درس علم داده با هدف به‌کارگیری روش‌های ارزیابی کلان داده‌ها در رشته مهندسی شیمی ایجاد شده است. در حالی که در بسیاری از دانشگاه‌های مطرح دنیا این رشته به‌عنوان یک رشته دانشگاهی تدریس می‌شود، تا کنون هیچ توافقی در رابطه با محتوای دوره یا سرفصل‌های آن ایجاد نشده و هر مؤسسه و دانشگاهی به سبک خود عمل می‌کند. اما هدف نهایی آن آشنایی دانشجویان کارشناسی با اصول ابزارها و روش‌های داده‌های بزرگ و به‌کارگیری این دانش در تحقیقات است. اگرچه مؤسسه‌های دانشگاهی شروع به تدوین برنامه‌های درسی علم داده کرده‌اند، اما هنوز پیشرفت‌های زیادی وجود دارد که باید صورت گیرد. فهرستی از دانشگاه‌های معتبر دنیا که مفاهیم علم داده را آموزش می‌دهند، در جدول ۳ آمده است.

جدول ۳. برنامه رشته مهندسی شیمی با مفاهیم علم داده در دانشگاه‌های معتبر دنیا

دانشگاه	عنوان درس - انگلیسی	عنوان درس - فارسی	جزئیات
دانشگاه فناوری ماساچوست ^۷	Engineering computation and data science ^۶	محاسبات مهندسی و علوم داده	<ul style="list-style-type: none"> - معرفی ابزارهای مدرن، الگوها و کتابخانه‌ها برای محاسبات توزیع‌شده غیرهم‌زمان شامل جداول درهم‌ساز^۱ توزیع‌شده، درختان مرکل^۲، و اثبات دانش صفر^۳ - آموزش پاک‌سازی^۴ و فیلتر کردن داده‌ها، رگرسیون خطی، و الگوریتم‌های اصلی یادگیری ماشین، مانند خوشه‌بندی^۵، طبقه‌بندی، درخت تصمیم‌گیری - مهندسی در محاسبات با تأکید بر علم داده - تمرین‌های کلاسی و تکالیف هفتگی به‌صورت گروهی
	Process data analytics ^۹	تجزیه و تحلیل داده‌های فرایند	<ul style="list-style-type: none"> - شامل مباحث یادگیری ماشین، داده‌های کلان، روش‌های بیزی^۸، طراحی آزمایش و تشخیص الگوهای تولید فرایند (به‌عنوان مثال پیش‌بینی خروجی، کنترل فرایند، شناسایی و تشخیص عیب‌ها) - تجزیه و تحلیل داده‌های فرایندی

1 - Hash tables

2 - Merkle trees

3 - Zero knowledge proofs

4 - Data cleaning

5 - Clustering

6 - <http://catalog.mit.edu/search/?P=1.00>, accessed Jun 2020.

7 - Massachusetts Institute of Technology

8 - Bayesian methods

9 - <http://student.mit.edu/catalog/m10a.html>, accessed Jun 2020.

ادامه جدول ۳

<ul style="list-style-type: none"> - درس اختیاری بین رشته‌های مهندسی شیمی و مواد - ارائه درس از سال ۲۰۱۴ - سه واحد درسی 	<ul style="list-style-type: none"> - علم داده و یادگیری ماشین در مهندسی شیمی و مواد 	Data science . & Mach. Learn. in Chem. & Mat. Engr ^۱	دانشگاه استنفورد ^۲
<ul style="list-style-type: none"> - بررسی جنبه‌های اجتماعی و اخلاقی علم داده - سیاست‌گذاری، مدیریت و تصمیم‌گیری - مقدمه عمومی علم داده - آمار و یادگیری ماشین - مصورسازی داده‌ها - مدیریت داده‌ها 	<ul style="list-style-type: none"> - علم داده 	Data science ^۳	دانشگاه واشنگتن ^۴
<ul style="list-style-type: none"> - انجام پروژه‌های گروهی بر روی چالش‌های مهندسی شیمی 	<ul style="list-style-type: none"> - چالش‌های علم داده برای سامانه‌های شیمیایی 	Challenges in data science for chemical systems ^۵	
<ul style="list-style-type: none"> - آموزش روش‌ها از طریق سخنرانی‌های کوتاه، فعالیت‌های درون کلاس و به‌کارگیری آن‌ها در علم داده به‌منظور تکمیل پروژه‌های متمرکز بر ارزیابی توصیفی و پیش‌بینی‌کننده فرایندهای شیمیایی - مباحث تحت پوشش عبارتند از: ارزیابی سری زمانی، پردازش داده‌ها، کاهش ابعاد، یادگیری با نظارت تقویت‌شده و بدون نظارت، درخت تصمیم‌گیری و خوشه‌بندی، رگرسیون چندمتغیره و مصورسازی - آموزش پایتون به‌عنوان زبان برنامه‌نویسی 	<ul style="list-style-type: none"> - علم داده برای سامانه‌های شیمیایی 	Data science for chemical systems ^۶	مؤسسه فناوری کالیفرنیا ^۷
<ul style="list-style-type: none"> - درک ابزارهای علم داده و روش‌های مدل‌سازی محاسبات - استفاده از روش‌های یادگیری ماشین و هوش مصنوعی 	<ul style="list-style-type: none"> - تجزیه و تحلیل داده‌های مهندسی شیمی 	Chemical engineering data analysis ^۸	دانشگاه کلمبیا ^۹
<ul style="list-style-type: none"> - مفاهیم یادگیری ماشین و ابزارهای ریاضی لازم برای درک آن‌ها 	<ul style="list-style-type: none"> - آشنایی با یادگیری ماشین 	Introduction to machine learning ^{۱۰}	
<ul style="list-style-type: none"> - مقدمه‌ای بر زبان برنامه‌نویسی R - آموزش یادگیری ماشین 	<ul style="list-style-type: none"> - مقدمه‌ای بر علم داده 	An introduction to data science ^{۱۱}	دانشگاه ملی سنگاپور ^{۱۲}
<ul style="list-style-type: none"> - ارائه مبانی ریاضی و فنون یادگیری ماشین - آموزش یادگیری ماشین و داده‌کاوی به‌منظور پیش‌بینی فرایندها 	<ul style="list-style-type: none"> - ریاضیات یادگیری ماشین 	Mathematics of machine learning ^{۱۳}	دانشگاه توکیو ^{۱۴}

1 - <https://stanford.app.box.com/s/euoq66bprz08cpawmoowyukj9naqjdf>, accessed Jun 2020.

2 - Stanford University

3 - <https://escience.washington.edu/data-science-courses-at-the-university-of-washington/>, accessed Jun 2020.

4 - University of Washington

5 - <https://www.cce.caltech.edu/academics/courses/che-142>

6 - California Institute of Technology

7 - <https://www.cce.caltech.edu/academics/courses/che-141>, accessed on June 2020.

8 - <https://cheme.columbia.edu/ms-chemical-engineering-concentration-data-and-computational-science>

9 - Columbia University

10 - <https://www.eng.nus.edu.sg/wp-content/uploads/sites/4/2020/05/Direct-Entry-Students-AY2020-2021.pdf>

11 - <https://catalog.he.u-tokyo.ac.jp/detail?code=FEN-CO2d21L1&year=2020>

12 - National University of Singapore

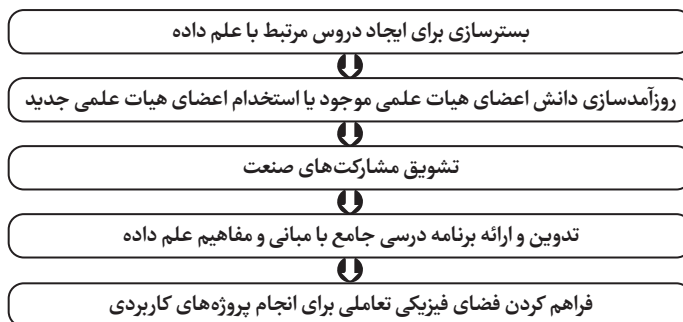
13 - <https://catalog.he.u-tokyo.ac.jp/detail?code=FEN-MP4d30L1&year=2020>

14 - University of Tokyo

۵. پیشنهاد ایجاد درس علم داده

تبدیل شدن به متخصص علوم داده به مجموعه‌ای از دانش‌ها و مهارت‌های مختلف نیاز دارد که مستلزم روشی معقول و مؤثر برای ادغام این مفاهیم و اصول با برنامه درسی دانشگاهی می‌باشد. مانند هر برنامه آموزشی جدید، برنامه درسی، منابع و دانشکده‌ها و برخی از ابزارهای ارزیابی باید مشخص و پیاده‌سازی شوند. با این حال، برنامه‌های علم داده با توجه به ماهیت بین‌رشته‌ای، گسترده بودن موضوع‌ها، دستیابی به اطلاعات و نیز زیرساخت‌های محاسباتی مورد نیاز در مقیاس بزرگ، چالش‌های خاصی را مطرح می‌کنند. بنابراین راه‌اندازی یک برنامه کارشناسی جدید در علم داده، ممکن است در بسیاری از مؤسسه‌ها اقدام مهمی باشد که مدیران و توسعه‌دهندگان برنامه‌ها را با تصمیم‌های بی‌شماری روبه‌رو خواهند کرد. محبوبیت دوره‌ها و برنامه‌های علم داده از چندین طریق بر زیرساخت‌های دانشگاهی تأثیر می‌گذارد. دانشکده‌ها و مدیران باید چگونگی تطابق اهداف آموزش علم داده با زیرساخت‌های فعلی مؤسسه را بررسی نمایند. از آنجایی که علم داده با ریاضیات، آمار، علوم رایانه و سایر حوزه‌ها مشترک است، مؤسسه‌ها باید در مورد ایجاد یک گروه جدید یا ادغام آن با سایر گروه‌ها تصمیم بگیرند.

همان‌طور که قبلاً ذکر شد، مفاهیم کلیدی مورد نیاز برای توسعه علم داده شامل مبانی ریاضیات، مبانی محاسباتی، مبانی آماری، مدیریت داده، توصیف و مصورسازی داده‌ها، مدل‌سازی داده‌ها و ارزیابی است. سپس این مهارت‌ها در طیف وسیعی از موقعیت‌های شغلی علم داده کاربرد خواهند داشت. با توجه به این که حوزه علم داده با سرعت زیادی در حال تکامل است، توجه به این نکته حائز اهمیت است که اغلب لازم است تا انواع شغل‌هایی که از علم داده استفاده می‌کنند و هم‌چنین مجموعه مهارت‌های علم داده برای دستیابی به موفقیت در آن مشاغل مجدداً مورد ارزیابی واقع شوند (National Academies of Sciences, 2018). چندین مانع برای راه‌اندازی و حفظ برنامه‌های علم داده وجود دارد و تا حدودی در مؤسسه‌های مختلف دانشگاهی برطرف شده است. شروع هر برنامه جدیدی که تقاضای زیادی داشته باشد، با چالش‌هایی همراه است (شکل ۶).



شکل ۶. مراحل ایجاد درس علم داده در دانشگاه‌ها

جدیدترین نمودار جامع درسی برنامه کارشناسی مهندسی شیمی در ایران، توسط دانشگاه امیرکبیر ارائه شده است که براساس آن می‌توان پیشنهادهایی برای تغییر برنامه درسی موجود و گنجاندن مفاهیم علم داده به دوره‌های موجود را از طریق جایگزینی سرفصل دروس اصلی یا انتخابی و یا برگزاری کارگاه‌ها اضافه کرد. پیشنهادهای به شرح زیر هستند:

- اصول محاسبات / روش‌های عددی در مهندسی شیمی: اضافه کردن روش‌های داده‌کاوی و یادگیری ماشین به سرفصل دروس
- دروس اختیاری: تعریف درس علم داده به‌عنوان درسی اختیاری
- کارگاه نرم‌افزار: آموزش زبان برنامه‌نویسی پایتون و R، آشنایی با ابزارهای مدیریت و ارزیابی و استخراج دانش از مجموعه داده‌های عظیم مکانی-زمانی
- مدل‌سازی ریاضی در مهندسی شیمی: آموزش مدل‌سازی و روش‌های ارزیابی داده‌ها و فرایندهای شیمیایی از طریق روش‌های علم داده، شامل پایگاه‌های داده، داده‌کاوی، یادگیری ماشین و مصورسازی داده‌ها

این روش‌ها برای دانشجویان و فارغ‌التحصیلان به‌عنوان یک مهارت مؤثر تلقی می‌شوند و آن‌ها را متناسب با نیازهای صنایع فرایندی امروز با درک دقیق مفاهیم و کاربردهای ارزیابی داده‌های بزرگ آشنا می‌سازند.

۱-۵. فهرست سرفصل‌ها و رئوس مطالب پیشنهادی برای درس "داده‌کاوی در مهندسی شیمی" براساس توضیح‌های ارائه شده در بالا، نویسندگان پیشنهاد می‌دهند که درس علم داده به‌عنوان یک درس اختیاری، مطابق با سرفصل‌های ذیل، به فهرست دروس دوره کارشناسی، کارشناسی ارشد یا دکتری اضافه گردد. بدیهی است با توجه به حجم بالای مطالب در این درس، می‌توان عناوین نام‌برده را در دو یا هر سه دوره تحصیلی تقسیم‌بندی نمود.

درس اول: مقدمه‌ای بر داده‌کاوی

درس دوم: مفاهیم و اصول پایه داده‌کاوی

درس سوم: روش‌های پیش‌پردازش داده‌ها (Pre-processing)

- پاک‌سازی (Data cleaning)
- تجمیع (Data integration)
- کاهش ابعاد (Dimensionality reduction)
- گسسته‌سازی (Discretization)
- روش تحلیل اجزا اساسی (Principal component analysis)
- روش تحلیل اجزا مستقل (Independent component analysis)

- روش حداقل مربعات جزئی (Partial least squares)
- روش‌های پیشرفته و ترکیبی
- **درس چهارم:** انبارش داده‌ها (Data warehousing)
- مدل‌سازی داده‌های چندبعدی
- معرفی فناوری تحلیل برخط اطلاعات
- روش‌های پیاده‌سازی سامانه‌های انبارش داده‌ها
- آشنایی با سامانه‌های ثبت داده‌ها در فرایندهای شیمیایی
- **درس پنجم:** روش‌های طبقه‌بندی (Classification) و پیش‌بینی داده‌ها (Prediction)
- معرفی مفاهیم پایه و شاخص‌های ارزیابی فرایند طبقه‌بندی
- معرفی روش‌های پایه طبقه‌بندی
- درخت تصمیم (Decision tree)
- طبقه‌بندی بیزین (Bayesian)
- طبقه‌بندی مبتنی بر قواعد (Rule-based)
- روش‌های تجمیع (Ensemble)
- روش نزدیک‌ترین همسایه‌ها (K-Nearest Neighbors) و روش یادگیرنده‌های تنبل (Lazy learners)
- شبکه‌های عصبی مصنوعی (Artificial Neural Networks)
- الگوریتم‌های ژنتیک (Genetic algorithms)
- روش‌های مبتنی بر منطق فازی (Fuzzy logic)
- روش‌های نیمه‌نظارتی (Semi-supervised)
- روش‌های شبکه‌های باور بیزین (Bayesian belief networks)
- ماشین‌های بردار پشتیبان (Support vector machines)
- **درس ششم:** روش‌های خوشه‌بندی (Clustering)
- معرفی مفاهیم پایه و شاخص‌های ارزیابی فرایند خوشه‌بندی
- معرفی روش‌های پایه خوشه‌بندی
- روش‌های مبتنی بر افزار (Partitioning)
- روش‌های سلسله‌مراتبی (Hierarchical)
- روش‌های مبتنی بر چگالی (Density-based)
- روش‌های مبتنی بر ساختار (Grid)
- خوشه‌بندی داده‌های با ابعاد زیاد (High-dimensional data)

- خوشه‌بندی مبتنی بر قیود (Constraint-based)
- خوشه‌بندی مبتنی بر مدل احتمالاتی
- خوشه‌بندی داده‌های گراف و شبکه
- درس هفتم: روش‌های تحلیل داده‌های پرت (Outlier analysis)
- معرفی مفاهیم و تعاریف پایه و نیز انواع داده‌های پرت
- بررسی روش‌های تشخیص و تحلیل داده‌های پرت
- روش‌های آماری
- روش‌های مبتنی بر مجاورت (Proximity)
- روش‌های مبتنی بر خوشه‌بندی
- روش‌های مبتنی بر طبقه‌بندی
- روش‌های تشخیص و تحلیل داده‌های پرت برای داده‌های با ابعاد زیاد (High-dimensional data)
- درس هشتم: تشخیص عیب، شناسایی عیب و تحلیل علیت در فرایندهای شیمیایی (Fault detection, Diagnosis and root cause analysis)
- روش‌های کاوش و ثبت داده‌ها
- بررسی روش‌های کاربردی در پیش‌پردازش
- بررسی روش‌های کاربردی در تشخیص و شناسایی عیب
- بررسی روش‌های تحلیل علیت (آنتروپی انتقال و تحلیل گرنجور)
- درس نهم: تشخیص و پیش‌بینی مدل‌ها و رفتار فرایندهای شیمیایی
- روش‌های کاوش و ثبت داده‌ها
- داده‌کاوی و طراحی و سنتز مواد
- داده‌کاوی و مدل‌سازی واکنش‌های شیمیایی
- داده‌کاوی و پدیده‌های انتقال
- داده‌کاوی و زیست‌فناوری^۱
- داده‌کاوی و ایمنی فرایندی

هم‌چنین شایان ذکر است که به منظور افزایش بهره‌وری این دوره، می‌توان آموزش یکی از نرم‌افزارها یا زبان‌های برنامه‌نویسی مهم داده‌کاوی را در کنار این درس مورد توجه قرار داد. پیشنهاد می‌گردد به فراخور نیاز و صلاحدید شورای آموزشی و تحصیلات تکمیلی دانشکده‌ها یا دانشگاه‌ها و یا اعضای هیئت علمی محترم، یکی از نرم‌افزارها و زبان‌های برنامه‌نویسی متلب، پایتون یا R آموزش داده شود تا به این شکل روش‌ها و الگوریتم‌های تدریس شده به شکل عملی مورد ارزیابی قرار گیرند.

۱- زیست‌فناوری برابر نهاد فرهنگستان برای بیوتکنولوژی است.

۶. نگاهی به آینده و جمع بندی

امروزه با پیشرفت فناوری، حجم داده‌های دیجیتال با سرعت زیادی در حال رشد است. داده‌های بزرگ شامل مجموعه‌های زیادی از داده تولیدشده از منابع ناهمگن و بسامد بالا هستند که می‌توانند با استفاده از روش‌های محاسباتی، پردازش و ارزیابی گردند. ظهور علم داده به‌عنوان یک حوزه علمی و دانشگاهی مرتبط با مفهوم داده‌های کلان از مجموعه داده‌هایی تشکیل شده و کاربردهای موفقیت‌آمیز زیادی از علم داده در صنعت، بخش مالی، مراقبت‌های بهداشتی، حمل‌ونقل، آموزش، خدمات حرفه‌ای و غیره گزارش شده است. بسیاری از بزرگ‌ترین اکتشاف‌ها و تصمیم‌های جهان در علم، فناوری و تجارت، اکنون براساس تجزیه و تحلیل مجموعه‌های عظیم داده‌ها انجام می‌شود. بنابراین مهندسان شیمی، همانند مهندسان رشته‌های دیگر، برای بهره‌گیری از داده‌هایی با افزایش حجم روزافزون نیازمند ابزارهای علم داده هستند. تصور کنید اکنون سال ۱۴۱۰ است و متولدین سال ۱۳۸۸ از دانشگاه فارغ‌التحصیل شده‌اند. بیش از ۳۰ سال است که میلیارد‌ها حسگر و دستگاه شروع به ذخیره مداوم داده‌ها در پایگاه‌های مختلف داده کرده‌اند. بسیاری از تحول‌های داده‌محور که پیش‌بینی آن‌ها دشوار بوده است، فراگیر و حائز اهمیت شده‌اند. بنابراین، دور از انتظار نیست که مؤسسه‌های دانشگاهی دنیای داده‌محور سال ۱۴۱۰ را تجسم کنند. در حالت ایده‌آل از منظر پیشرفت آینده علم داده، تمامی صنایع خصوصی و دولتی می‌توانند به‌طور مؤثر و کارآمد از داده‌ها استفاده کنند. مشاغل علم داده فراوان خواهد بود. لذا روش‌های علم داده باید توسط برنامه‌ها و دروس علم داده به‌طور مداوم آموزش داده شوند تا نیازهای دانشجویان و نیروی کار را تحقق بخشد. در این صورت جای شک نیست که برای انتقال از دنیای سال ۱۳۹۹ به دنیای احتمالی سال ۱۴۱۰، چارچوب آموزشی کارشناسی باید تغییر کند. امروزه از یک مهندس شیمی در سطح دانشگاهی یا صنعتی، بیش از پیش خواسته می‌شود که مجموعه داده‌های پیچیده و بزرگ را ذخیره‌سازی و ارزیابی کرده و قادر به مصورسازی نتایج نیز باشند. از این‌رو بازمینی برنامه‌های آموزشی رشته مهندسی شیمی در دوره کارشناسی با دیدگاه تقویت آموزش علم داده حائز اهمیت ویژه‌ای است تا بتوان جایگاه مناسبی را برای این آموزش تعریف کرد. با تقویت و بازنگری در برنامه‌های آموزشی، دانشجویان رشته مهندسی شیمی می‌توانند با برخی از فناوری‌های علم داده در قالب دروس آشنا گردند. لازم است که برنامه‌های درسی مهندسی شیمی، شامل آموزش در علوم داده، آمار، یادگیری ماشین و برنامه‌نویسی رایانه باشد تا دانش‌آموختگان آمادگی بهتری برای غلبه بر چالش‌ها در آینده داشته باشند. این تحقیق، ضمن بیان مفاهیم و ضرورت استفاده از ارزیابی داده‌های کلان و هم‌چنین بررسی رویکردهای مؤسسه‌های دانشگاهی، دیده وسیعی برای کمک به هدایت این تحول ارائه کرده است. لازم به ذکر است که با رشد فناوری، این بینش باید به‌طور مداوم تکامل یابد و پالایش شود.

References:

- 42 Years of Microprocessor Trend Data. (2020). <https://www.karlsruhp.net/2018/02/42-years-of-microprocessor-trend-data/>.
- Ardeshiri Lajimi, K., & Abbasnejad, R., (2016), Big data from yesterday to tomorrow, *2nd International Conference on Research in Science and Engineering*, Istanbul, Turkey. https://www.civilica.com/Paper-ICRSIE02-ICRSIE02_118.html
- Asadi Samadi, D., Keramati Thani, S., Amini, A., and Rahimi, H. (2018). Security challenges of 5V principles in bigdata big data, national congress of fundamental research in computer engineering and information technology, Tehran – Shahid University Beheshti, *Permanent Secretariat of the Conference*. https://www.civilica.com/Paper-COMCO05-COMCO05_060.html
- Akragholipour, M., Khodavardi, F., and Safa, R. (2016). A study of the relationship between cloud computing, big data and internet of things, *Third National Conference on Engineering Science Development, Mazandaran-Tonekabon*, Institute of Higher Education. https://www.civilica.com/Paper-AIHE10-AIHE10_040.html.
- Abonyi, J., Kulcsar, T., Balaton, M., & Nagy, L. (2013). Historical process data based energy monitoring-model based time-series segmentation to determine target values. *Chemical Engineering Transactions*. (pp. 26-35).
- Abonyi, J., Farsang, B., & Kulcsar, T. (2014). Data-driven development and maintenance of soft-sensors. *In 2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMII)* (pp. 239-244). IEEE.
- Ansari, A., Mohaghegh, S. D., Shahnam, M., & Dietiker, J. F. (2019). Modeling average pressure and volume fraction of a fluidized bed using data-driven smart proxy. *Fluids*, 4(3), (pp. 119-123).
- Ayers R. (2017). Four things to know about big data in chemical engineering. <https://www.aiche.org/chenected/2017/08/four-things-know-about-big-data-chemical-engineering>.
- Beck, D. A., Carothers, J. M., Subramanian, V. R., & Pfaendtner, J. (2016). Data science: Accelerating innovation and discovery in chemical engineering. *AICHE Journal*, 62(5), (pp.1402-1416).
- Beck D., Pfaendtner J., Carothers J., Subramanian V. (2017). Data science for chemical engineers, *AICHE Journal, Chemical Engineering Progress*, 113(2), (pp. 21-26).
- Cao, L. (2018). Data science discipline. *In Data Science Thinking* (pp. 129-160). Springer, Cham.
- CCAC (Community College of Allegheny County). (2020). *Data Analytics Technology (788.1): Associate of Science*. https://www.cca.edu/academics/_academics-files/instructions-fall2019.pdf.
- Chiang, L. H., & Colegrove, L. F. (2007). Industrial implementation of on-line multivariate quality control. *Chemometrics and Intelligent Laboratory Systems*, 88(2), (pp. 143-153).
- Chiang, L., Lu, B., & Castillo, I. (2017). Big data analytics in chemical engineering. *Annual Review of Chemical and Biomolecular Engineering*, 8, (pp. 63-85).
- Cozad, A., Sahinidis, N. V., & Miller, D. C. (2014). Learning surrogate models for simulation based optimization. *AICHE Journal*, 60(6), (pp. 2211-2227).
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), (pp. 64-73).
- Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), (pp. 745-766).
- Duever, T. A. (2019). Data science in the chemical engineering curriculum. *Processes*, 7(11), 830.
- Earnshaw, R., Dill, J., & Kasik, D. (2019). Data science and visual computing. *Springer International Publishing*.
- Guan, Y. (2017). Application of data mining in chemical production. *Chemical Engineering Transactions*, 62, (pp. 805-810).
- Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, 2, (pp. 652-687).
- Kano, M., & Fujiwara, K. (2012). Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications. *Journal of Chemical Engineering of Japan*, 12, (pp. 156-167).
- Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.

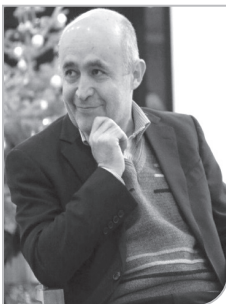
- Kheradmandi Nia, Sh., & Sotoudeh Qarabagh, R., (2019). Additional training for chemical engineers from the perspective of consulting engineers. *Iranian Journal of Engineering Education*, 20 (77), (pp. 1-17).
- Ma, Y., Niu, P., Yan, S., & Li, G. (2018). A modified online sequential extreme learning machine for building circulation fluidized bed boiler's NOx emission model. *Applied Mathematics and Computation*, 334, (pp. 214-226).
- Marr, B. (2017). The complete beginner's guide to big data everyone can understand. <https://www.forbes.com/sites/bernardmarr/2017/03/14/the-complete-beginners-guide-to-big-data-in-2017/#3d6175f7365>.
- Montgomery College. (2020). DATA 101 - Introduction to data science, https://catalog.montgomerycollege.edu/preview_course_nopop.php?catoid=8&coid=11413.
- Nashua Community College. (2020). Why choose foundations in data analytics? <http://www.nashuacc.edu/academics/associate-degrees/stem-and-advancedmanufacturing/398-foundations-in-data-analytics>.
- National Academies of Sciences, Engineering, and Medicine. (2018). Data science for undergraduates: Opportunities and options. National Academies Press.
- Piccione, P. M. (2019). Realistic interplays between data science and chemical engineering in the first quarter of the 21st century: Facts and a vision. *Chemical Engineering Research and Design*, 147, (pp. 668-675).
- Reinsel, D., Gantz, J., & Rydning, J. (2017). Data age 2025: The evolution of data to life-critical. Don't focus on big data, <https://assets.ey.com/content/dam/ey-sites/ey-com/en-gl/topics/workforce/Seagate-WP-DataAge2025-March-2017.pdf>.
- Sun, Y., Yan, H., Zhang, J., Xia, Y., Wang, S., Bie, R., & Tian, Y. (2014). Organizing and querying the big sensing data with event-linked network in the internet of things. *International Journal of Distributed Sensor Networks*, 10(8), (pp. 218-221).
- Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers & Industrial Engineering*, 115, 319-330.
- University of California, Berkeley. (2020). Principles and techniques of data science, <http://www.ds100.org/sp20/>.
- Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003). A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3), (pp. 293-311).
- Venkatasubramanian, V. (2009). Drowning in data: informatics and modeling challenges in a data rich networked world. *AIChE Journal*, 55(1), (pp. 2-8).
- White, D. (2016). Big Data: What is it? CEP Mag. 112, (pp. 33-35).
- Ziaei-Halimejani, H., Zarghami, R., Mansouri, S. S., Mostoufi, N. (2021). Data-driven fault diagnosis of chemical processes based on recurrence plots. *Ind. Eng. Chem. Res.* 60, 3038-55.



◀ **دکتر میترا سادات سادات لواسانی:** مدرک کارشناسی خود را در رشته مهندسی شیمی از دانشگاه صنعت نفت، مدرک کارشناسی ارشد را از دانشگاه فردوسی مشهد و مدرک دکتری را از دانشگاه دانشگاه سیستان و بلوچستان اخذ کرده است. تمرکز تحقیقاتی ایشان روی شبیه سازی دینامیک سیالات محاسباتی و ساخت برج تقطیر دیوار میانی در مقیاس پیلوت و آنالیز داده می باشد.



◀ **مهندس هومن ضیائی حلیمه جانی:** مدرک کارشناسی و کارشناسی ارشد خود را از دانشگاه تهران در رشته مهندسی شیمی اخذ کرده است و در حال حاضر دانشجوی دکتری مهندسی شیمی در دانشگاه تهران می باشد. تمرکز تحقیقاتی ایشان بر روی کاربرد علوم داده در مهندسی شیمی و تحلیل داده های کلان می باشد.



◀ **دکتر رحمت ستوده قره باغ:** مدرک دکترای خود را در مهندسی سیال سازی از دانشگاه پلی تکنیک مونترال کانادا اخذ کرده اند. ایشان از سال ۱۳۷۸ فعالیت خود را به عنوان عضو هیأت علمی در دانشگاه تهران آغاز و هم اکنون نیز به عنوان استاد در دانشکده فنی مشغول به فعالیت هستند. نامبرده در کنار فعالیت های آموزشی و پژوهشی در زمینه مهندسی فرآیند و آموزش مهندسی، مشغول به کار اجرایی و مدیریتی هم هستند. دکتر ستوده برنده سال ۹۴ جایزه علامه طباطبایی از بنیاد ملی نخبگان بوده و بیش از ۲۵۰ مقاله علمی در نشریات بین المللی، ملی و کنفرانس ها، شش کتاب و چهار فصل کتاب تألیف نموده اند. ایشان به عنوان عضو کانون کارشناسان رسمی دادگستری در رشته صنایع نفت فعالیت می نمایند.



◀ **دکتر رضا ضرغامی:** مدرک دکترای خود را در مهندسی سیال سازی از دانشگاه تهران اخذ کرده اند. ایشان در حال حاضر به عنوان استاد و رئیس دانشکده مهندسی شیمی در پردیس دانشکده های فنی مشغول به فعالیت هستند. فعالیت های آموزشی و پژوهشی ایشان در زمینه دینامیک و کنترل فرآیند، دینامیک سیالات محاسباتی و تحلیل داده های کلان می باشد. دکتر ضرغامی به عنوان پژوهشگر نمونه دانشگاه، بیش از ۲۰۰ مقاله علمی در نشریات بین المللی، ملی و کنفرانس ها، یک کتاب و چندین فصل کتاب تألیف نموده اند. مسئولیت سردبیری نشریه انگلیسی مهندسی شیمی و نفت دانشکده مهندسی شیمی به عهده ایشان می باشد.



◀ **دکتر نوید مستوفی:** مدرک دکترای خود را در مهندسی سیال سازی از دانشگاه پلی تکنیک مونترال کشور کانادا اخذ کرده اند. ایشان از سال ۱۳۷۹ فعالیت خود را به عنوان عضو هیأت علمی در دانشگاه تهران آغاز کرده و هم اکنون نیز به عنوان استاد تمام در دانشکده فنی مشغول به فعالیت هستند. ایشان هم چنین در کنار فعالیت های آموزشی و پژوهشی در در زمینه مدلسازی و مهندسی فرآیند، دینامیک سیالات محاسباتی، یک دوره مسئولیت ریاست دانشکده مهندسی شیمی را به عهده داشتند. دکتر مستوفی پژوهشگر برجسته دانشگاه تهران بوده و بیش از ۳۶۰ مقاله علمی در نشریات بین المللی، ملی و کنفرانس، هفت کتاب و چندین فصل کتاب تألیف نموده اند.