

名称：MANAGEMENT of DATA (MOD) セミナー ～ソーシャル・データサイエンスとAIソリューション～

概要：データサイエンス・エンジニアリング（DS&E）の世界は、ChatGPTの登場以来、その適用領域と多様性が爆発的に広がっていますが、さまざまな技法・ツールが登場する中、どれが生き残ってゆく（定着する）のか誰にも分からない状況にあります。このような中では、専門家に丸投げしたり、あるいは、流行りを追いかけたりすることになりがちですが、AI革命を生きるには、すぐに陳腐化する表層ではなく中身＝本質を理解することや、どこから手をつけどう学んでゆくと良いか土地勘を持つこと、が重要です。具体的には、DS&Eの背後にある理論や処理プロセスを理解すること、誤りなく使えること、結果を正しく解釈できること、何が出来・何が出来ないかを知ること、そして、何が新規で意義深いかを判断できること、が必要となります。本講義では、データ分析から社会実装までの基層的なデータ・リテラシーを、大学院レベルで、短期間・集中的に講義します。これにより、受講者が受講後、AI社会を生きる自分なりの術を構築してゆくことが期待されます。

実施概要

日程：2025年8月～2025年9月の水曜日と日曜日、全10回（日程の詳細は下記の表をご覧ください）

時間：水曜日は18:30-21:00（150分、休憩10分を含む）、土曜日は9:30-12:00（150分、休憩10分を含む）

場所：東京科学 キャンパス・イノベーションセンター

実施形態：

- a. 全員、ネットワーク接続できるPCの持参をお願いします。OSはWINDOWS 10以降、/MacOS 10.13 High Sierra以降をお願いします。また、情報共有のためzoomを使いますが、ハウリング防止のためヘッドセットをご用意ください。
- b. 対面での授業とグループワークを行います（対面参加をお勧めしますが、やむを得ずオンライン参加される場合は、学習効率が下がることをご承知おきください）。
- c. 講義室に来られなくなった場合は、早めにご連絡お願い致します。
- d. Google Colaboratoryを使用します（<https://colab.research.google.com/?hl=ja>）. 使ったことがない方は、予習ビデオを準備しますので、視聴の上、ご参加ください。
- e. Python未学習の方は基礎クラスからの参加をお勧めします。この場合、予習ビデオの他に簡単なドリルも準備しますので、基礎クラス開始前にビデオの視聴及びドリルをこなした上で、ご参加ください。
- f. ソリューション構築クラス・総合演習クラスでは生成AIを使用します。ChatGPT Plusの使用料の自己負担をお願いします（税抜20\$/月）。登録方法は、ソリューション構築クラス開始前にお知らせします。

対象：新学習指導要領を境にデータ・リテラシーを自然に身に付けた世代が現れ、今後、社会が急激に変わってゆくことから、現役世代のリカレント教育が急がれます。そこで、データサイエンスに関して教養以上の理解を得たい方、統計の学習に留まらない実践的なデータ分析・ソリューション構築に興味のある方、大学院（修士課程、博士後期課程）入学希望の方、参加者同士のネットワークを構築したい方、などを対象とします。

講師：中井豊（芝浦工業大学名誉教授、東京工業大学非常勤講師）、三好昭生（株式会社インサイト代表取締役）、淀薫（株式会社インサイト 技術開発部主査）、中村伸也（株式会社インサイト技術開発部）

コーディネーター：猪原健弘（東京工業大学リベラルアーツ研究教育院教授）

表（各コースの番号、履修クラス、日程、受講料） ※ やむを得ず休講が生じた場合の代替日：9/27土

コース	番号	各クラス・各コースの授業日程（2025年8月～2025年9月の水曜日と日曜日）					受講料 （税込）
		①基礎クラス （Python速習）	②機械学習クラス	③テキスト分析 クラス	④ソリューション 構築クラス	⑤総合演習クラス	
総合コース	MOD-2025-S	8/20水, 23土	8/27水, 30土	9/3水, 6土	9/10水, 13土	9/17水, 20土	110,000円
アレンジ履修1	MOD-2025-1	12/16土, 17日	12/23土, 24日	－	－	－	44,000円
アレンジ履修2	MOD-2025-2	12/16土, 17日	－	1/6土, 7日	－	－	44,000円
アレンジ履修3	MOD-2025-3	－	12/23土, 24日	1/6土, 7日	－	－	44,000円
アレンジ履修4	MOD-2025-4	12/16土, 17日	12/23土, 24日	1/6土, 7日	－	－	66,000円

① 基礎クラス（含 Python速習）

内容：課題の解決に当たっては、問題の構造を明らかにし解決の鍵（KFS: Key Factor for Success）を見出す必要があります（データサイエンス）。そして次に、データ分析に留まらず、分析結果を踏まえたソリューション構築を行わなければなりません。本クラスでは、データを用いた科学的課題解決（MANAGEMENT of DATA）のコンセプトを学ぶとともに、教材データ（医療サービスデータ）を用いて、基礎的なスキルであるデータの前処理や記述統計を行います。また、データサイエンスへの理解が表層的に止まらないよう、自ら手を動かしデータを処理する機会を提供することで、履修者の理解を深めます。具体的には、演習問題（Python基礎ドリル）を使い、データサイエンスの代表的プログラミング言語であるPythonを速習し、基礎的なデータ分析のスキルを身に付けます。

各回の概要：

DAY-1（中井）150分（休憩10分を含む）

- ・ガイダンス、チーム分け、MOD概論（データサイエンスとソリューション）
- ・Python基礎（変数、データの読み書き、IF文/FOR文など）

DAY-2（中井）150分（休憩10分を含む）

- ・データ前処理（欠損・外れ値処理、型変換、標準化）
- ・記述統計（散布図、相関行列）
- ・課題説明：診療科を選んで分析

② 機械学習クラス

内容：本クラスでは、解決の鍵を科学的に探索・検証するため、定量データに対する機械学習を学びます。具体的には、機械学習の流れを学ぶとともに、教材データ（医療サービスデータ）を用いて、代表的な機械学習である重回帰分析、ロジスティック回帰分析、主成分分析、クラスター分析、決定木、深層学習を、自らデータを扱うことで学びます。

各回の概要：

DAY-1（中井）150分（休憩10分を含む）

- ・機械学習の流れ：前処理（One hot encoding）、訓練・検証（層化サンプリング）、汎化性能・過学習
- ・教師あり学習：重回帰・ロジスティック回帰分析、結果の解釈（混同行列・統計的有意）、変数の選択（多重共線性・Lasso回帰）

DAY-2（中井）150分（休憩10分を含む）

- ・教師なし学習（主成分分析・クラスター分析、結果の解釈）
- ・発展的話題：決定木、深層学習
- ・課題説明：診療科を選んで分析

③ テキスト分析クラス

内容：言うまでもなく数字だけでは問題の本質を見誤る恐れがあります。現場の感覚、中でも人々の想いや意見を理解することは極めて重要です。この意味でネット上に溢れるコメントは人々の本音を照らし出す貴重な情報源です。本クラスでは、問題の構造を意味的に検証するため、テキストデータに対する分析の進め方を学びます。具体的には、スクレーピングやAPIを使ってネットからのテキストデータの取得を行うとともに、取得したテキストデータの前処理（形態素解析、分散表現）を学びます。そして、教材データ（医療サービスデータ）を用いて、代表的なテキストマイニングである共起ネットワークやトピック分析を、自らテキストデータを扱うことで学びます。

各回の概要：

DAY-1（中井）150分（休憩10分を含む）

- ・スクレーピング（Beautifulsoup）、API、形態素解析（Mecab）
- ・分散表現（辞書、BOW、ストップワード）

DAY-2（中井）150分（休憩10分を含む）

- ・記述統計（単語の出現頻度、TF-IDF、共起ネットワーク）
- ・テキストの意味の計量（文章の類似度、トピック分析）
- ・課題説明：診療科を選んで分析

④ ソリューション構築クラス

内容：ビジネス課題の解決において、データサイエンスと生成AIの統合的活用が不可欠となっています。本クラスでは、特に異常検知という普遍的な課題に焦点を当て、その理論的背景と実践的な解決手法を学びます。まず生成AIの基礎技術を一般的な文書処理の例を通じて学んだ上で、製造ラインでの品質管理やECサイトでの異常購買検知など、実践的な異常検知の応用例を取り上げます。データ分析の結果を生成AIで効果的に活用し、実践的なソリューションを構築するための方法論を、技術的な背景の理解に重点を置いて学習します。

各回の概要：

DAY-1（三好）150分（休憩10分を含む）*座学セッション

- ・生成AIの技術的基礎
 - トランスフォーマーの数理解説 →具体例：文章要約タスクでの注意の動作解説
 - 高度化技術の理論 →具体例：法律文書での文脈維持の仕組み
 - 実用上の制約と対策 →具体例：文書生成での一貫性維持の課題
- ・プロンプトエンジニアリングの実践
 - データ分析結果の定式化手法 →具体例：数値データからの報告文生成
 - コンテキスト設計と管理 →具体例：多段階指示での精度向上事例
- ・ワークフロー構築の基礎
 - システム設計原則
 - Difyを用いたパイプライン構築 →具体例：文書処理システムの設計事例

DAY-2（三好）150分（休憩10分を含む）*座学セッション

- ・異常検知の理論と応用
 - 統計的基礎
 - * 管理図法の数理 →具体例：ペットボトル充填量の管理図分析
 - * 移動平均モデルの理論 →具体例：お菓子の色むらトレンド分析
 - 機械学習アプローチ
 - * Isolation Forestの原理 →具体例：ECサイトでの異常購買検知
 - * オートエンコーダーの理論 →具体例：製造ライン複合センサーデータの異常検知
- ・生成AIとの統合実践
 - 異常検知結果の解釈支援 →具体例：充填ライン異常の自動解釈生成
 - アラートシステムの設計 →具体例：色検査異常の警報文自動生成
 - 応用事例による解説
 - * 製造ライン品質管理の総合事例
 - * EC購買行動分析の総合事例

⑤ 総合演習クラス

内容：データサイエンスと生成AIの統合的活用をテーマに、実践的な試行錯誤を通じて総合演習を行います。異常検知を題材に、生成AIとの対話による実装、結果の評価、改善というサイクルを実体験します。特に、プロンプトエンジニアリングの実践と、異常検知アルゴリズムの実装を組み合わせることで、より深い理解と実践力を養います。この演習で得られる知見は、データ分析と生成AIの統合という観点で、幅広い業務領域（製造、人事、顧客管理など）に応用可能です。

各回の概要：

DAY-1（三好・中村（・淀：⑤総合演習クラスを受講生が8名を超えた場合に参加））150分（休憩10分を含む）*演習セッション

- ・異常検知タスクの導入
 - 課題設定の説明
 - 評価基準の確認

（次ページから続く）

- ・実装フェーズ 前半
 - 生成AIとの対話による実装（各自）
 - コード生成と実行の試行錯誤
 - 中間経過の共有
- ・実装フェーズ 後半
 - グループでの改善検討
 - 改良版の実装と検証
 - 成果の整理

DAY-2（三好・中村（・淀：⑤総合演習クラスの受講生が8名を超えた場合に参加））150分（休憩10分を含む）*演習セッション

- ・ソリューション完成
 - 前回の課題対応
 - 実装の最終調整
 - 発表資料作成
- ・グループによる成果発表
 - グループ発表
 - 総合討論：実務応用の可能性

備考：

④ソリューション構築クラスの開始前に、Pythonの基礎を中心にそれまでのクラスに関する質疑応答（オンライン）を行う日時を設けます。すべての受講者が対象です。参加は任意です。