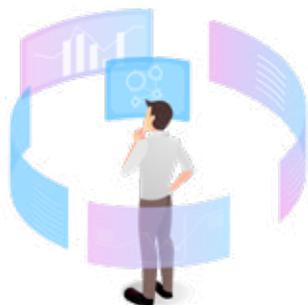


データ分析者育成プログラム

Associate コース

実際のデータを分析し活用するために、
統計学の手法、代表的な機械学習の手法（分類・回帰）を学び、
目的・用途に応じた機械学習モデルの構築ができるレベルを目指します。



Point- 1 講義要素を盛り込んだ python 演習

自らデータ加工や分析をするための知識を習得でき、
python コードを用いてすぐに実践できます。

Point -2 講義終了後もストリーミング配信で何度でも見直せる！

各プログラム毎に収録した動画をストリーミング配信します。
何度でも視聴できますので復習に最適です。

Point-3 プログラムの追加も可能！

プログラムの内容はご要望に応じてカスタマイズが可能です。

※本カリキュラムは一般社団法人データサイエンティスト協会
「スキルチェックリスト 2021 年版」 Associate Data Scientist (独り立ちレベル) を
参考に構成しています

ご準備いただくもの

- Assistant コースの内容を理解していることが望ましいです。
- PC (タブレット PC 以外の PC をご用意ください)
講義の進行には Zoom、Slack、Google Colaboratory を使用します。
セキュリティ上ご利用が難しい場合はご相談ください。

基本カリキュラム

開催形態：オンライン (Zoom) ※対面をご希望の場合、ご相談に応じます

講義時間：90分

開催期間：ご要望に応じます

No.	プログラム名	習得する知識やスキル	主な項目
1	記述統計・推測統計	実際のデータを活用した統計量（標本平均・標本分散）の算出および標本平均・標本分散と母平均・母分散の関係について習得する	母集団・母平均・母分散、標本平均・標本分散、推定量、不偏推定量、描画（ヒストグラム、箱ひげ図、散布図）、標本分布（t分布、区間推定、有意性検定）、ベイズ統計
2	クラスタリング	教師なし学習の1種であるクラスタリングについて理解を深めるとともに、ある特徴量空間上のデータを複数のクラスに分ける手法を習得する	クラスタリングを行う目的の確認、階層的クラスタリング・非階層的クラスタリング、類似度の算出（距離の定義）、手法（階層：ワード法等、非階層：K-means法等）、分析結果の数値の解釈・次元の呪い
3	相関分析	相関分析の手法と理論を学び、相関分析データの正しい読み方を習得する	データの可視化（散布図・ヒストグラムなど）、近似曲線の作成、相関係数の算出と検証、相関と因果の関係性、外れ値の影響
4	決定木分析（分類）	決定木を用いた分類を行い、交差検証法を用いたモデル評価およびモデル選択について習得する	データクレンジング、訓練データとテストデータへの分割、決定木の構築と予測（ジニ不純度）、汎化性能の推定（k-分割交差検証法）、最適なパラメータの探索
5	回帰分析（回帰）	回帰分析の手法を学び、多重共線性に留意しつつ、予測モデルを構築するとともに、決定係数、t値、p値等によりモデルの精度を評価する	単回帰分析・重回帰分析、平均二乗誤差・二乗平均平方根誤差・決定係数、多重共線性・VIF（分散拡大係数）、情報量基準（AIC・BIC）、検定、交差検証法、過学習、汎化性能、正則化
6	ロジスティック回帰分析（分類）	多変量解析の一種であり、ある事象の発生率を判別するロジスティック回帰分析を理解するとともに、モデル構築から評価までを習得する	ロジスティック回帰分析と重回帰分析の違い、偏回帰係数、オッズ比、シグモイド関数、最小二乗法・最尤法、メリット・デメリット・留意点 適合度評価（p値、AIC、BIC）
7	主成分分析	多くの変数を持つデータを集約して主成分を作成する統計的分析手法である主成分分析について理解し、その手法を習得する	多数の量的な説明変数の1～3次元への要約、固有値、寄与率、累積寄与率、主成分負荷量、主成分得点、変数間の関係性の可視化 注意点（各主成分の意味の考察、寄与率が低い等）
8	次元削減	多次元のデータをなるべく情報を失わないように低次元のデータに落とし込む次元削減について理解を深め、その手法を習得する	データ圧縮、ノイズ除去・特徴抽出、データ可視化、主成分分析（PCA）、多次元尺度構成法（MDS）、正準相関分析（CCA）、グラフ埋め込み（GE）
9	アンサンブル学習	データ数が少ない場合の手法である決定木の予測精度を改善する手法であるアンサンブル学習の手法について習得する	決定木の復習、バギング、ランダムフォレスト、アダブースト、勾配ブースティング
10	特徴量エンジニアリング	機械学習モデルの構築に欠かせない特徴量エンジニアリングの手法を習得し、データ整備ができるようになる	特徴量エンジニアリングの目的・ドメイン知識、新たな特徴量の追加、集計方法、特徴量の変換、外れ値・異常値・欠損値・特徴量の取舍選択、クラスタリングで特徴量作成、主成分分析で次元圧縮、

この講座についてのお問い合わせ

株式会社 Tokyo Tech Innovation（東工大 100%出資子会社）

データ分析者育成プログラム担当

<https://tokyotech-i.co.jp/>

〒152-8550 東京都目黒区大岡山 2-12-1 T-3（東工大 蔵前会館 3F）

TEL 03-6425-7305 FAX 03-6425-7306

mail: info@tokyotech-i.co.jp（右の QR コードもご利用ください）

営業時間：8:30～17:30（土日祝を除く）



Tokyo Tech Innovation
公式サイト内
お問い合わせページ



Tokyo Tech Innovation
公式サイト内
詳細ページ