
前言

無論是否意識到，機器學習（ML）都已經成為日常生活中不可分割的一部分。每次進入像是 YouTube 和 Amazon.com 之類的網站時，你都在和驅動人性化推薦的 ML 產生互動；這表示，網站上的產品顯示方式取決於 ML 演算法對你品味和興趣的判斷。不僅如此，還有基於 ML 的評論審核系統，可以標記垃圾郵件或惡意評論、評論審核等等，在像 YouTube 這樣的網站上，還會有 ML 生成的字幕和翻譯。

除了購物和娛樂，ML 也滲透到我們生活中的各個層面；例如，當你在線上轉帳時，ML 演算法會檢查這筆交易是否存在詐騙風險。我們就是生活在一個以資料和 ML 演算法為基礎而建構的軟體時代。

這些軟體都需要專業人才的設計和建構，這也提升對軟體技能的需求，使 ML 職業的地位不斷提升，薪資也水漲船高。這些因素只是 ML 職業吸引力的一部分，打造我們生活中不可或缺的產品和功能更是主因，因為 ML 技術推動人工智慧的進展，所以這個討論同樣也適用於「人工智慧職業」。

不過，要踏入 ML 領域極富挑戰性。ML 領域的職位通常要求較高學歷文憑，2010 年代大多數的職位需要有博士學位。儘管從 2010 年代末期開始，招募啟事上對所要求的學歷已經逐漸下降，但我在網路上仍然經常看到至少具備碩士學位以上的建議；但就算是那些具備足夠學歷的人，要在資料和 ML 領域中找到工作也很不容易。這到底是網路上提供了錯誤的建議，還是這建議過於籠統和含糊其辭？

我參加過許多 ML 職位的面試，從初階層級、資深層級、主任 +¹ 和首席層級² 上都曾經順利通過面試，在整個過程中，我親身經歷了雄心壯志的應徵者在 ML 面試期間遭遇的相同困難和挫折，我也曾送出無數履歷表卻沒有收到任何回應，在電話面試篩選中就被刷掉過，在等待回應時飽嚙焦慮不安的滋味；甚至專程從多倫多飛到舊金山現場面試，但還是沒通過。我曾經應徵過資料科學家和機器學習工程師（MLE）的職位，但當面試官似乎更想尋找資料工程師或資料分析師時，也很讓我感到困惑不解。

除了作為面試者的經驗以外，我也累積好幾年身為面試官的經驗。作為在 ML 領域工作的一部分，我審查和過濾好幾百份履歷表，主持過多次面試，並參加過許多決策委員會；作為技術領導階層，即兩間公司的首席層級一部分，我審查過工作描述，面試過建教合作生、實習生、初階層級應徵者，以及資深和主任層級的受僱者。我在本書中包含工作應徵者所會犯下的錯誤提示，而正是這些錯誤導致面試官決定不讓他們進入下一輪面試。經常會聽到這句話：「如果應徵者不這樣就好了，他們本來很有希望的。」這本書將幫助你避免犯下這類明顯錯誤。

事實上，對於求職者面試中有許多不言而喻的標準。例如，良好的溝通和團隊合作技巧，可能就不會很明顯地寫進對工作內容的描述中；這些期望並不是惡意地要從工作描述中省略，而是因為業內人士將其視為最低要求。我最近看到一些大公司的 ML 招募啟事明確地將「溝通技巧」列在他們要求清單的最前面，以試圖改善工作描述的明確性。

除了對新求職者和有經驗求職者一視同仁的這些潛在期望以外，面試過程在不同職位和不同公司之間的極大差異，也可能令人感到困惑。甚至連多年在 Google 從事資料工作的作家 Randy Au 在出於好奇，而查看當下資料科學家和 ML 職位招募啟事之後，也指出「情況……各有不同」³。

1 主任 + (Staff+) 指的是高於資深層級的職位。

2 在技術領域中工作層級的推進通常是初階 / 中階層級 → 資深層級 → 主任層級 → 首席層級，然而不同公司之間會有細微差異。例如，有些公司會合併主任和首席層級。

3 Randy Au, 〈Old Dog Revisits the DS Job Market out of Curiosity〉, 「Counting Stuff」(部落格), 2022 年 12 月 1 日, <https://oreil.ly/yzIsx>。

許多人都希望能有個路徑圖，或關於邁入 ML 領域完整循序漸進的保證。例如，最好在大學中主修什麼科系、在哪些地方實習？什麼是最好的業餘專案？應該掌握哪些 Python 函式庫？這些我都能夠理解，對於工作面試過程中的每一個步驟，為了獲悉盡可能多的資訊，我請教過很多朋友關於這方面的事情。我為了面試後是否應該發送後續行動的電子郵件而煩惱，而且查了很多線上論壇，想知道要不要發送這封信；這樣會打擾到面試官嗎？還是他們也期望我這麼做？連這麼小的事情都讓我嚴重焦慮，而且我希望能得到一個明確的答案，而不是「這要視情況而定」，或「應該不會造成什麼傷害」。這是一本我希望當時的我在碰到這些問題時都能拿來參考的書！

現在，我已經是位於另一邊的面試官了，我了解成為雇主傾向僱用求職者的各種情況；現在，對於過去的許多問題，我也有了具體答案，而且對於邁入 ML 領域的路徑圖也有了更多親身體驗；但儘管有這樣具品質保證的路徑圖，它也不會如你所想。在我了解 ML 和資料科學領域的時候，我老早就選好大學主修科目，也畢業了，而且正在攻讀經濟學碩士學位。在大學期間我沒有任何實習經驗；相反地，空閒時間我都在設計、玩電動遊戲，或是到處交朋友。事實上，ML 工作的路徑圖非常有彈性，而且就算起步較晚，也沒什麼大不了，沒有什麼事情會真的來不及。

當我找第一個 ML 工作的時候，我並沒有完成所有一定要完成的事情，但無論如何，我都以沒有實習過的學生身分通過工作面試。關於面試過程，我知道的可能比很多人少，也沒有做到所有該做的事，卻仍然在 ML 領域中茁壯成長，這也是為什麼，我反而能用這樣的觀點來寫作。實際上，沒有什麼是該做的事，只有適合你的事情。

我不會告訴你這樣的事情，「只要在大學主修某個科系，然後在某公司實習過，就萬無一失了。」我想要的是，為每一種不同類型的人寫本專書。一體適用、規範性的路徑圖，只要碰到不在路徑圖的狀況就會失靈；但如果你學會在不完全依賴路徑圖的情況下，也知道應對方法，這樣無論發生什麼事，你都能夠打造自己的路徑圖。

本書將說明成為領航者的方法，並打造自己的路徑圖，無論你是否主修 STEM⁴、是否有實習經驗、有相關工作經驗、有 ML 工作經驗，或具備 ML 以外的工作經驗等，只要能夠堅持，就算你主修的不是一般推薦的領域也沒問題。如果認為以前的工作經驗和 ML 沒有直接關係，那也不要緊；我將指導你增強和利用過去經驗，以及獲得其他相關經驗的辦法。

我主張應該依據自己的情況活化、量身訂做職業路徑圖，因為在我自己的職業生涯中，就曾經碰過很多不在單一路徑圖的情況：

- 以經濟學碩士身分，在大型上市公司⁵找到初階層級的資料科學家（ML）工作
- 在新創公司找到較為資深的職位，這間公司在我剛加入時大約有 200 名員工，全盛時大約有 400 名員工
- 在新興中大型上市公司找到首席資料科學家的工作

根據產業、公司規模、ML 團隊大小和公司生命週期的階段（例如新創階段），雇主對於我所能掌握的事情會有不同期望。如果我只遵照線上建議，或參加過使用不同工作面試流程公司面試者的建議，那我可能無法通過面試；不，絕對沒辦法。為了能夠順利通過面試，我每次都必須準備不同面試的方式才能成功。透過所有我個人經驗和參加過（說真的）數百次 ML 面試，對於在 ML 和資料科學工作面試中得分並成為成功應徵者而言，我已經找到了成功的模式；透過經驗和學到的教訓，現在才有可能寫這本書來幫助懷有野心的應徵者。

成功的求職者知道面試過程中的每個步驟，都是試圖評估各個場景下的事情；但不幸的是，只靠簡單地露個面和具備技術技能並不一定足夠。就像學校中的考試一樣，仔細查看過教學大綱並了解所有考試範圍的人，更有成功的可能。在這種情況下，對於要應徵的每項工作，你都應該試著對它的教學大綱實施逆向工程。

隨著我在 ML 領域中獲得越來越多的經驗，從雄心壯志應徵者那裡也看出越來越多的問題，我參加過很多次咖啡聊天，到目前為止已經超過 100 次了，而且為了幫助更多的人，多年來我一直在部落格「susanshu.com」撰寫有關職業生涯的指引。如果這本書的問世能有機會幫助到更多人，我的決定就顯而易見了。

4 科學、技術、工程和數學。

5 上市公司（*public company*）指這間公司有公開交易的股票。

為什麼要選擇機器學習工作？

我曾經說過，無論了不了解或喜不喜歡 ML，它在我們的日常生活中都已經非常普遍，你可能在自己的生活經歷上也感受到，所以才會讓你產生好奇並拿起這本書！我也會概述我的經驗，這可能會增強你的動機，或讓 ML 領域更具吸引力，而讓你注意到它。

作為一位從事技術工作的人，我認為 ML 是開發可以影響數百萬使用者高價值產品的重要領域；在我離開學校後的第一份工作，就有機會從事這樣的專案，我認為如果當時的我對機器學習不嫻熟，在職業生涯如此早期，可能也不會有這樣的責任和機會。

依我看來，ML 是一個有趣而且能夠實現個人抱負的領域。我喜歡學習新的技術和研究，如果你認同的話，你也會喜歡從事 ML 領域這方面的工作。但是在這個領域中的快節奏創新也有它負面影響；例如，想要專注在家庭或生活中其他重要層面的時候，還要持續追上新進展可能會讓人精疲力盡。如今，即使我非常專注於在週末參加社交活動，或寫本書之類的其他活動上，但只要不會耗費到太多時間，我仍然會抓住機會學習。我也會在工作期間花些時間聽聽網路上的講座或看看書。這不是 ML 所獨有的，但很多人都說過，持續學習 ML 的步調，會比其他需要學習新觀點的技術相關工作再快一些。

當然，還有薪資方面的考量；一般而言，ML 工作的報酬很好，我不但可以養活自己，甚至還達成許多能改善我和家人生活的財務目標，能夠做到這些，都要感謝我的 ML 職業生涯。另一方面，我能夠有那麼多的成就，也都是因為 ML 領域和社群所促成：我經常飛到世界各地在研討會上演講，且研討會多到我已經在為數年後做準備；還能會晤在 ML 這麼棒地方工作的卓越人才，並親眼目睹 ML 和 AI 領域的進展，這都是能在這個行業工作的好處。

無論拿起這本書的動機為何，為了讓你能夠順利通過 ML 的工作面試，並克服中間過程的障礙，我都希望能夠與你分享一些讓人成功的技巧和工具。

這本書將幫助你了解以下內容：

- 各種類型的 ML 職位，並找出你最可能通過面試的職位
- ML 面試的構成要素

- 找出你在技能上的缺口，好有效地鎖定面試準備
- 順利的通過技術和行為面試

我也會添加我在 O'Reilly 講授的線上直播訓練課程經常被問到的問題，可以把它視為和我一起參加的咖啡聊天，以及從以下不同來源所獲得具支持性的見解：

- 可能不是非「典型」教育或職業背景的應徵者，順利通過面試的辦法
- 大幅增加履歷表得到初篩認可的機會
- 資深以及更高職位的 ML 面試進行方式

當然還有其他項目。

本書受眾讀者

在深入探討各章內容之前，我想先概述以下也許能引起你共鳴的場景，這些是我寫這本書主要想訴求的讀者群：

- 渴望成為業界 ML / AI 從業人員的應屆畢業生。
- 正在轉換成每天以 ML 工作為重心職位的軟體工程師、資料分析師，或其他技術 / 資料專業人員。
- 具有其他領域經驗，且有興趣轉換到 ML 領域的專業人員。
- 正在重回面試戰場，並且以轉換到不同職位或晉升和負更大責任為目標，而且希望能夠全方位複習 ML 內容的有經驗資料科學家或 ML 從業者。

如果以下情況是形容你，那你也可以從這本書獲益：

- 想要獲得主持 ML 面試靈感的經理，或是不想在零散的線上資源浪費太多時間，才能獲得面試過程概觀的非技術人員。
- 具有 Python 程式設計和 ML 理論基本知識，而且很好奇 ML 領域能不能成為未來職業選擇，並有心探討的讀者。

本書總覽

本章致力於幫助你區別各種 ML 職位，並詳細說明整個面試過程，如圖 1-1 所示：

- 工作應徵和履歷（第 2 章）
- 技術面試
 - 機器學習（第 3、4 和 6 章）
 - 編碼 / 程式設計（第 5 章）
- 行為面試（第 7 章）
- 你的面試路徑圖（第 8 章）
- 面試後和後續動作（第 9 章）

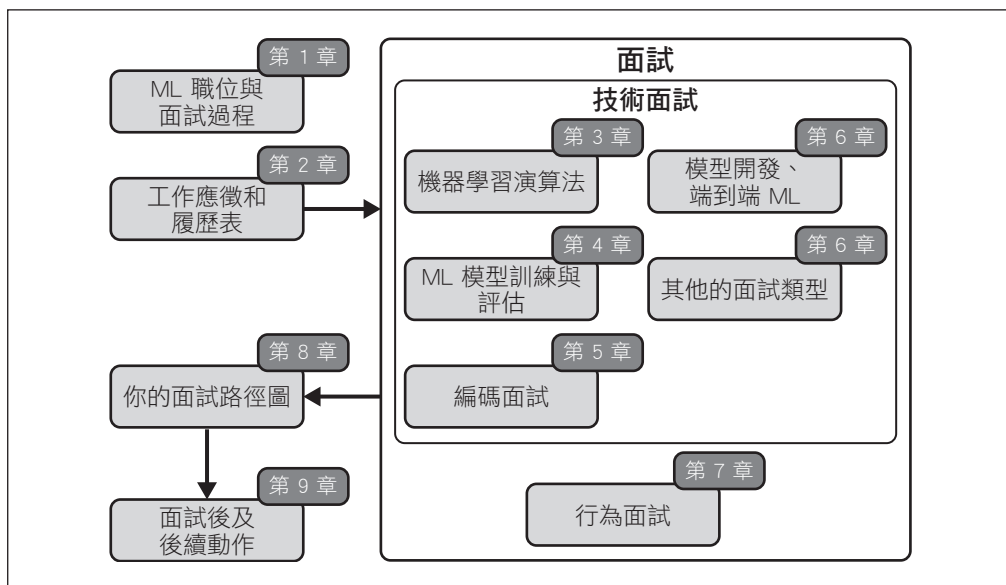


圖 1-1 全書各章總覽，以及它們與 ML 面試過程的關聯。

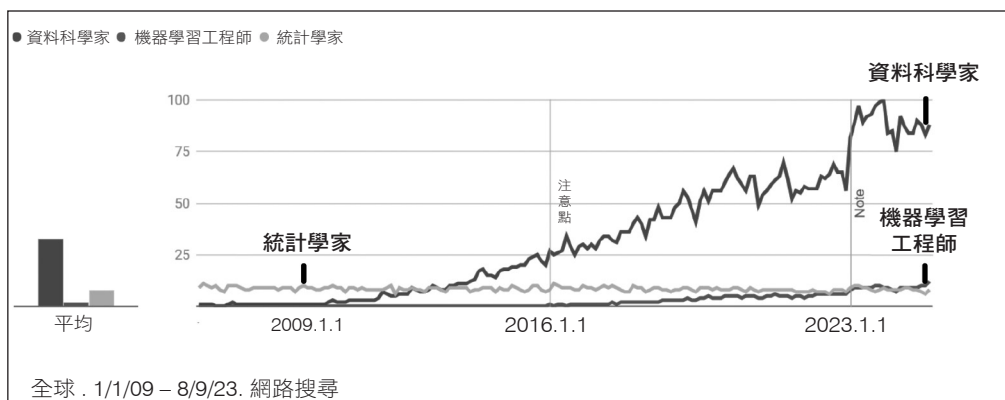


圖 1-2 「資料科學家」、「機器學習工程師」和「統計學家」在 Google Trends 上搜尋的熱度（2023 年 8 月 9 日檢索）。

隨著社群媒體、網路推薦系統和其他現代使用的案例增加，一些公司開始蒐集更細膩的資料，像是點擊流資料（*clickstream data*），也就是蒐集使用者瀏覽網站或應用程式時的資料；另一個最近的進步，是一般公司能夠儲存來自機器和物聯網（IoT）設備大量的遙測資料。從前，資料科學家可能要處理每週或每日更新的資料；現在，隨著許多應用程式更頻繁或即時的更新，需要更多基礎架構為網路產品和應用中的 ML 功能提供服務，因此圍繞著這些功能，也造就了更多的工作機會。

簡而言之：當機器學習生命週期（*machine learning lifecycle*）發展得更加複雜，也會產生更多工作職稱，以闡述完整 ML 團隊現在需要的新技能。本章稍後會更詳盡的說明工作職稱和 ML 生命週期。

所有這些都發生在過去十年之內，但公司並不總是能立即改變工作職稱以反映更為專業的職位。無論如何，作為一個求職者，了解這段歷史，在應徵各家公司，發現完全相同職稱卻有不同工作內容時，能有效減少所造成的困惑和挫折。對 ML 相關工作職稱先前的趨勢可參考表 1-1，ML 相關工作職稱的目前趨勢，則可參考表 1-2。

表 1-1 ML 和資料工作職稱先前的趨勢

ML 和資料工作職稱	工作職稱先前的趨勢
資料科學家	什麼事都要做
資料分析師	專門負責與業務決策相關的資料分析

機器學習職位的三大支柱

為了幫本書其餘部分做好準備，這裡會複習一下我稱為 ML 和資料科學職位的三大支柱：

- 機器學習演算法與資料直覺
- 程式設計與軟體工程技能
- 執行與溝通技巧

這些是在 ML 工作面試時會用來評估你技能的廣泛類別。本書的重點集中在幫助你了解這些技能，並找出你目前經驗和技能，與這三個支柱下的經驗和技能之間的差異（參考圖 1-6）；後續章節會一一說明這些技能。

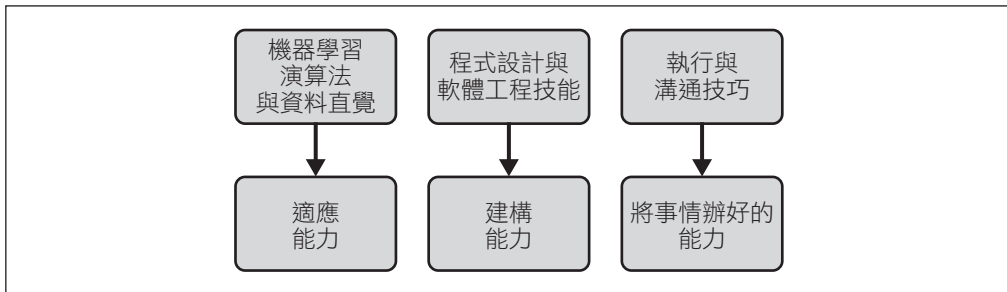


圖 1-6 機器學習工作的三大支柱。

機器學習演算法與資料直覺：適應能力

指的是你了解 ML 演算法和統計理論的基本工作原理，以及它們各自的權衡取捨；面臨真實世界 ML 專案工作中的開放式問題時，這是不可或缺的技能，沒辦法再像交學校作業那樣的按部就班。

具有資料直覺意味著當你面對新的問題時，會知道使用資料來解決它的方法；而且遇到新的資料或資料來源時，也會知道深入以評估它們的辦法。會先自問，這些資料適合 ML 嗎？適合什麼類型的 ML 模型？在將資料用到 ML 之前，這資料是否有任何問題？你會知道要問些什麼以及找出解答。

在機器學習工作面試過程中，各種不同類型的面試和面試問題，目的皆是在評估應徵者對這個支柱的知識與準備情況，第 3 章和第 4 章會再談到。

程式設計與軟體工程：建構能力

在從事一個專案的時候，具有交付所需要的程式設計技能，像是使用 Python 或用內部的部署過程處理資料，以便其他團隊可以使用來自 ML 模型的結果。

即使你很了解理論，但沒有程式設計或軟體工程¹⁴的意識，也無法使 ML 憑空出現。你需要用程式碼將資料和 ML 演算法連結，而演算法也要靠程式碼實現；也就是說，你必須將理論轉換成實踐。

對其他程式設計技能來說，ML 職位有高度需求的是（軟體）工程師從原型轉變到生產的能力，即整合和發布 ML 的能力。有些職位負責端到端的 ML：從研究和訓練模型到部署和生產；有些如 MLOps 工程師的 ML 職位，則負責建構能夠處理大量資料，以便能在幾秒甚至幾毫秒內，將 ML 回應給使用者需求的軟體基礎架構。

在 ML 面試過程中，各種不同類型的面試和問題都會評估應徵者在這個支柱上的技能，第 5 章和第 6 章會更仔細說明。

執行與溝通：在團隊中將事情辦好的能力

你能夠和不同職位的人一起工作；在 ML 中，我們和軟體工程師、資料工程師、產品經理和許多其他同事一起工作。在團隊中將事情辦好的能力包括一些軟技能，如溝通和一些專案管理技能。

例如，無法和團隊成員溝通對專案來說就是個阻礙（blocker）¹⁵，而且可能會導致你的 ML 專案停滯不前，或甚至惡化。即使你只和一個人，比如說老闆一起工作，你仍然需要在專案上提出報告，這也不乏溝通技巧。因此，在 ML 領域，高度需要的技能是能夠和非技術利害關係者溝通技術概念。

14 在處理設備上或邊緣 ML 更專業的職位中，擁有一些硬體基本知識也會有很大關係。

15 指阻止另一件事情，通常指專案或時間軸發生的商業用語。

你也需要一些專案管理的技能，好讓工作步入正軌。我們都在教育或自學的過程中學習管理待辦事項清單和行事曆的方法，但是因為現在你的專案行事曆取決於其他人的行事曆和優先事項，所以會更為混亂。即使你有專案和 / 或產品經理來讓團隊步入正軌，在某種程度上仍然需要管理自己。

沒有軟技能，就無法把事情辦好，這是不言而喻的事。不要成為那種只專注在技術技能，而忽略在面試中建立和展示軟技能的應徵者；第 7 章會深入研究 ML 面試在這個支柱上評估應徵者細節的方式。

三大 ML 支柱明確的最低要求

增強這三大 ML 支柱技能是一項艱鉅的任務，而且初階層級的職位通常只會期望你在每個支柱上有最低程度的要求，像是 3/10，如圖 1-7 所示。例如，一個接觸過程式設計的求職者，即使他們還不嫻熟或經驗不足，也可以經由訓練而改善。理想情況下，為了從其他求職者中脫穎而出，你至少要在與特定 ML 職位最相關的一個支柱上，表現得更為擅長，如在程式設計上為 5/10。

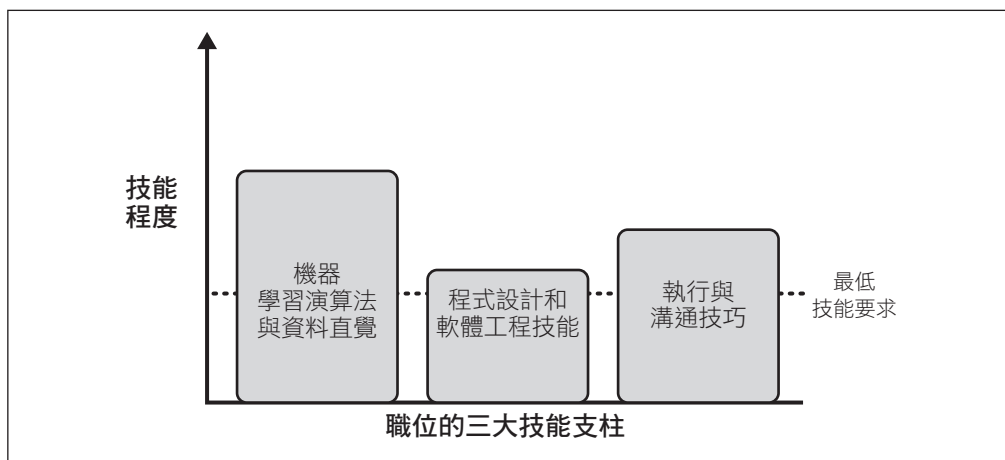


圖 1-7 對 ML 工作要求的最低技能程度（範例）。

對於資深的職位，這些基本最低要求會更高，但類似經驗法則仍然適用：有明確的最低技能要求，也因此，會根據職位將你擅長的技能拿來和其他應徵者比較。只訓練 ML 模型但不部署的資料科學家，對程式設計技能發展的要求，可能比不上 ML 理論和溝通技巧。

對於初階層級職位，的確比較不要求溝通支柱，但也不是 0/10 好嘛！因為要提高這技能，需要和一大群人合作，包括非技術同事在內等得來不易的經驗。這也提供某些求職者在這個支柱上的優勢：對於那些非傳統背景的應徵者，像是自學，或從軟體工程師職位及其他領域轉換過來的應徵者，擅長講述作品集發想過程和展示能力，都可以讓他們顯得比其他應徵者更優秀。

在總覽過這三個支柱後，你就可以利用這種心智模型脫穎而出。

機器學習技能矩陣

恭喜！這個部分相當龐大複雜，但你已經走到最後了！現在你檢視過機器學習生命週期和 ML 技能三大支柱的總覽，是時候將興趣和技能映射到工作職稱上了。

表 1-3 將大略提供想在特定職位上獲得成功，所需要學習的各項技能。範圍從 1 到 3 顆星，1 顆星代表重要性較低的技能，而 3 顆星代表非常重要的技能。

表 1-3 機器學習和資料技能矩陣

技能	工作職稱				
	資料科學家 (DS)	ML 工程師 (MLE)	MLOps 工程師	資料 工程師	資料 分析師
資料視覺化、溝通	★★★	★★	★	★	★★★
資料探索、淨化、直覺	★★★	★★★	★	★★★	★★★
ML 理論、統計	★★★	★★★	★★	★	★
程式設計工具 (Python、SQL)	★★★	★★★	★★★	★★★	★
軟體基礎架構 (Docker、Kubernetes、CI/CD)	★	★ 到 ★★★	★★★	★	★



在後續幾個章節會經常提到表 1-3，現在就存下來或夾上書籤！

看一看這些技能，你可以大致地將它們映射到前一節 ML 技能的三個支柱上，如表 1-4 所示。

表 1-4 機器學習和資料技能映射到 ML 工作的三支柱

支柱	ML 和資料技能
支柱 1 機器學習演算法與資料直覺：適應能力	資料探索、淨化、直覺 機器學習理論、統計 資料視覺化
支柱 2 程式設計與軟體工程技能：建構能力	程式設計工具（Python、SQL） 軟體基礎架構
支柱 3 執行與溝通技巧：在團隊中將事情辦好的能力	溝通等

如果你現在還無法完全確定每種類型的技能需求，也沒關係；第 2 章會重提這個矩陣，而且還有更詳細描述和自我評估檢核表。

給應屆畢業生的提醒

不需要擔心各項單獨技能，很多事情都是進了公司之後才會學到，但表現出你容易接受訓練，並且學習力強，絕對可以從其他求職者中脫穎而出。在 ML 職業生涯早期證明這一點的一個簡單方法，是對你目前尚無經驗的主題有高精度接觸，不夠深入也沒關係。例如，即使你不太常使用版本控制，但熟悉它將會是額外優點；你可以觀看一些可能只有 30 分鐘的影片，並在專案中花個 1 小時安裝 / 測試，而做到這點。

現在，將這一切結合在一起，在審視過機器學習生命週期（圖 1-5）以及機器學習技能矩陣（表 1-3）後，就可以來看看現在最適合你應徵的工作或獲得的技能！先將所有內容關連到 ML 和資料工作職稱目前的趨勢上（表 1-2），如圖 1-8 所示。

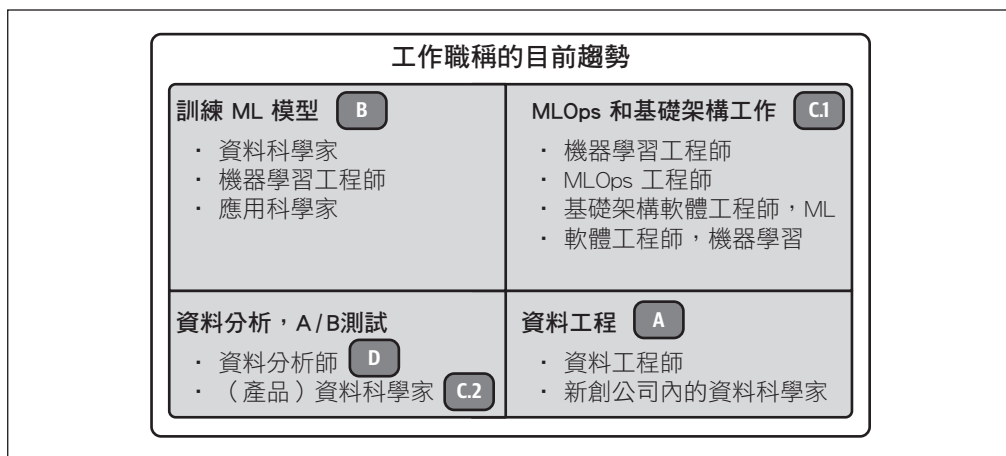


圖 1-8 常見的 ML 工作職稱，以及它們所對應的 ML 生命週期。

圖 1-8 中的字母標註可以映射到圖 1-5 中的標註，為方便而顯示如下：

- (A) 資料
- (B) 機器學習開發
- (C.1) ML / 軟體基礎架構
- (C.2) ML 假設測試 / 監控
- (D) 報告和儀表板



在後續幾章會一直提到圖 1-8，現在就儲存下來或夾上書籤！

當看到工作職稱並檢查招募啟事的詳細資訊時，你可以將它對應到這個職位日常可能會負責的工作。此外，依據你感興趣機器學習生命週期的部分，更可以準備和定位你想要的工作申請，這樣就不會意外的找錯了工作。