



《机器学习工程师行业调查》

|  |  |
| --- | --- |
| **专 业** | **计算机科学与技术** |
| **姓 名** | **王寄哲** |
| **日 期** | **2022/8/27** |

**摘 要**

本文对机器学习工程师这一行业做了详细的分析。从行业概述、就业情况对行业进行了概括性的描述。又列举了该行业内不同的工作方向进行了岗位职能介绍，并根据岗位职能提出了行业需要的主要能力与技术栈。接下来以两个大厂的招聘信息为例，分析了专业知识在行业中的实践路径，最后给出了就业准备建议。

关键词：机器学习 深度学习 数据挖掘 大数据分析

**目 录**

[第1章 行业概述与应届生就业现状 - 1 -](#_Toc14478)

[1.1 行业概述 - 1 -](#_Toc8865)

[1.1.1 应用领域 - 1 -](#_Toc23216)

[1.1.2 行业近况 - 1 -](#_Toc23216)

[1.1.2.1 发展速度 - 1 -](#_Toc22557)

[1.1.2.2 投资情况 - 1 -](#_Toc22557)

[1.2 应届生就业现状 - 2 -](#_Toc8865)

[第2章 岗位职能介绍 - 4 -](#_Toc14478)

[第3章 行业需要的主要能力与技术栈 - 5 -](#_Toc14478)

[3.1 主要能力 - 5 -](#_Toc8865)

[3.2 技术栈 - 5 -](#_Toc8865)

[第4章 专业知识在行业中的实践路径 - 6 -](#_Toc14478)

[第5章 就业准备建议 - 7 -](#_Toc14478)

[结 论 - 8 -](#_Toc5560)

**第1章** **行业概述与应届生就业现状**

## 1.1 行业概述

### 1.1.1 应用领域

机器学习指专门研究计算机怎样模拟或实现人类的学习行为以获取新的知识或技能的学科，使计算机重新组织已有的知识结构并不断改善自身的性能。机器学习以数据为基础，通过研究样本数据寻找规律，并根据所得规律对未来数据进行预测。机器学习是人工智能的核心，广泛应用于数据挖掘、计算机视觉、自然语言处理、生物特征识别等人工智能领域。

随着各行业对数据的需求量增多，对处理和分析数据的效率要求变高，一系列机器学习算法应运而生。机器学习算法主要是指通过数学及统计方法求解最优化问题的步骤和过程。针对不同的数据和不同模型需求，选择和使用适当的的机器学习算法可以更高效地解决一些实际问题。

机器学习不仅在基于知识的系统中得到应用，而且在自然语言理解、非单调推理、机器视觉、模式识别等许多领域也得到了广泛应用。一个系统是否具有学习能力已成为是否具有“智能”的一个标志。

### 1.1.2 行业近况

1.1.2.1 发展速度

2006年以来，深度学习的推广应用速度逐步加快，应用领域不断拓宽，其逐渐成为机器学习领域的主流算法。在深度学习的带动发展下，机器学习在人工智能应用市场中的应用占比从2014年的9.6%上升至2018年的12.2%。人工智能市场规模从2014年71.7亿元增长至2018年的339.0亿元，年复合增长率达47.5%，在人工智能商业化应用步伐逐步加快以及机器学习应用占比不断提高的发展背景下，机器学习发展进一步加快。机器学习广泛应用于金融、教育、医疗、工业、零售、能源等多个垂直领域，2014至2018年，各垂直领域的机器学习应用项目平均市场价格处于500万元至580万元区间内，根据各垂直领域机器学习项目平均市场价格以及项目数量进行计算，中国机器学习行业市场规模从2014年的8.7亿元增长至2018年的52.5亿元，年复合增长率达56.7%。随着机器学习算法的不断改进，其在各垂直领域的应用将进一步加深，各垂直领域的机器学习应用项目数量将不断增多，预计至2023年，机器学习市场规模将达336.7亿元，2019至2023年年复合增长率达46.3%。

1.1.2.2 投资情况

在近年获投的机器学习创业公司中，热门赛道集中于金融、互联网、工业、政务、医疗等。其中，金融赛道与机器学习契合度高且需求强烈，90%以上的机器学习企业都开展了金融相关业务板块，机器学习在金融领域的应用场景主要在智能风控、保险核定、精准营销等方面;机器学习在工业(含电力)领域也有着充分地施展空间，科学的算法模型应用能够帮助工业企业优化生产工艺、提升生产效率、减少资产损失;医疗领域，机器学习的应用集中于两方面，一是药物发现中通过算法提高靶点筛选、晶型预测等环节的效率，二是以算法模型赋能基因测序，提升疾病风险预测与辅助诊疗的准确性。

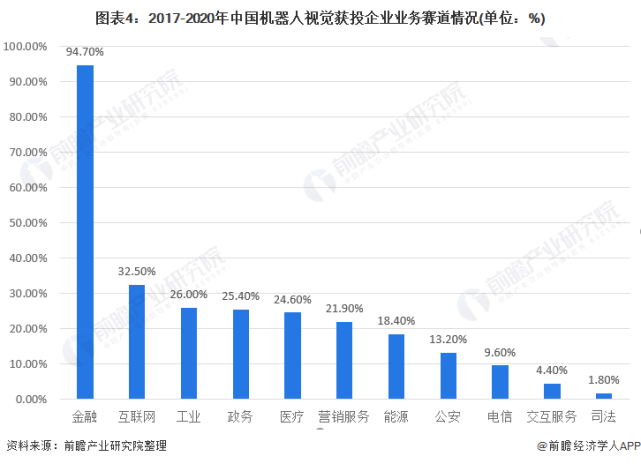


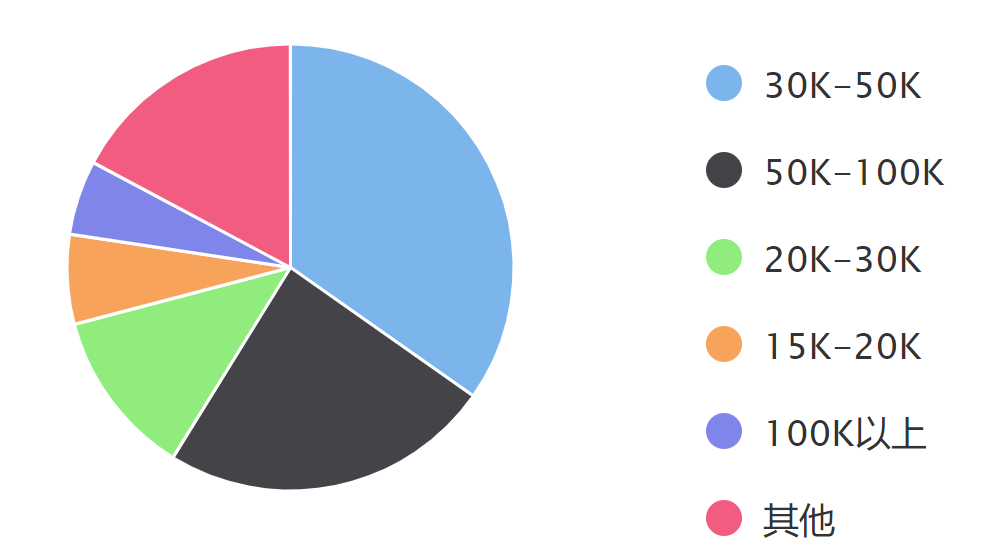
图1 2017-2020年中国机器人视觉获投企业业务赛道情况

## 1.2 应届生就业现状

根据某招聘网站信息：2021年：职位4.0K个，2021年较2020年增长了24%，2020年：职位 3.2K个，较去年下降了57%，，2019年：职位 7.5K个，较去年下降了20%，，2018年：职位 9.3K个，较去年增长了22%，，2017年：职位 7.6K个，较去年增长了423%

该职位招聘在30K-50K薪资范围占34.0%，该职位招聘在50K-100K薪资范围占24.0%，该职位招聘在20K-30K薪资范围占12.0%，该职位招聘在15K-20K薪资范围占6.0%，该职位招聘在100K以上薪资范围占5.0%，该职位招聘在其他薪资范围占17.0%。

表2 机器学习工程师月薪分布图



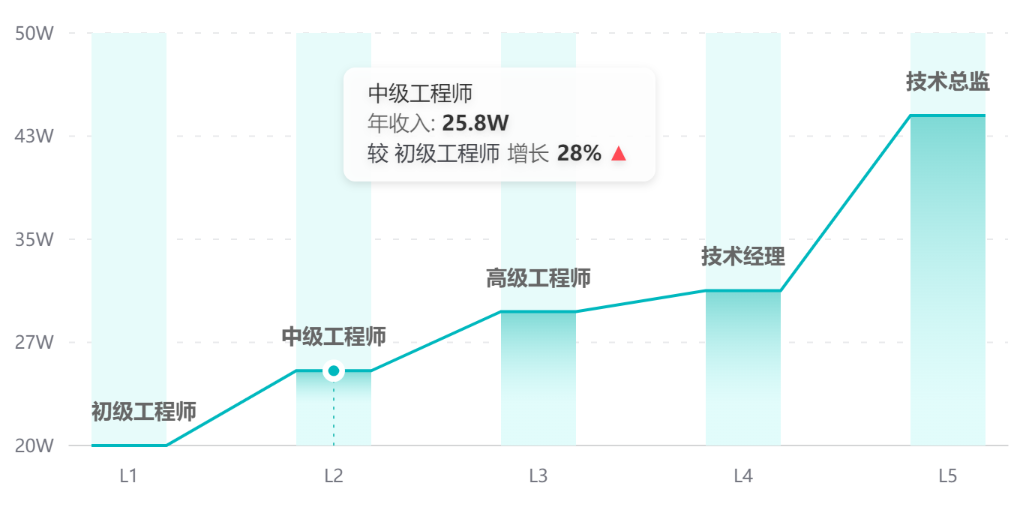


图2 就业发展

**第2章 岗位职能介绍**

由于机器学习并不针对于某一特定行业，而是可以结合并应用于各行各业，所以适合机器学习工程师的岗位种类非常丰富，跨度也非常大，因此所对应的职能也五花八门。这也就要求了机器学习工程师一定要具备一定的自学能力以及创新能力。下面介绍一些比较主流、热门的职位供参考：

推荐算法工程师：

运用各类机器学习模型，进行图像、语音、自然语言处理、搜索推荐等领域的算法设计、评估、优化，提供个性化推荐服务的工程师。

搜索算法工程师：

利用机器学习算法对搜索索引中的海量数据进行整理和排序，为用户提供最相关、最实用搜索结果（文字、图片、音视频、地图等）的计算机工程师。

深度学习算法工程师：

应用深度学习算法技术，解决各行业中的实际业务问题，探索人工智能与机器学习领域前沿技术进展的高级工程师。

智能驾驶系统工程师：

负责汽车智能驾驶系统架构设计的专业技术人员。

语音识别工程师：

针对具体应用场景，开发、测试和优化语音识别系统或工具，实现语音识别、语音指令、人机交互等功能的计算机工程师。

数据挖掘师：

运用机器学习技术处理海量数据，发现富有价值的业务信息的工程技术专业人员。

**第3章 行业需要的主要能力与技术栈**

## 3.1 主要能力

了解机器学习、深度学习、强化学习、迁移学习、主动学习、特征提取与稀疏学习、等级学习等相关知识。

负责机器学习、深度学习领域的技术研发工作，包括但不限于神经元网络模型设计与优化、强化学习、迁移学习、主动学习、维度降低、核方法、谱方法、特征提取与稀疏学习、等级学习、推荐、随机优化等的算法和系统研发等。

负责机器学习尤其是深度学习前沿问题的探索与研究，结合未来实际应用场景，提供全面的技术解决方案。

负责提供分布式的算法实现的解决方案，大幅提升算法计算规模和性能。

负责提供大数据分析建模方案，沉淀行业解决方案，协助拓展业务边界。

## 3.2 技术栈



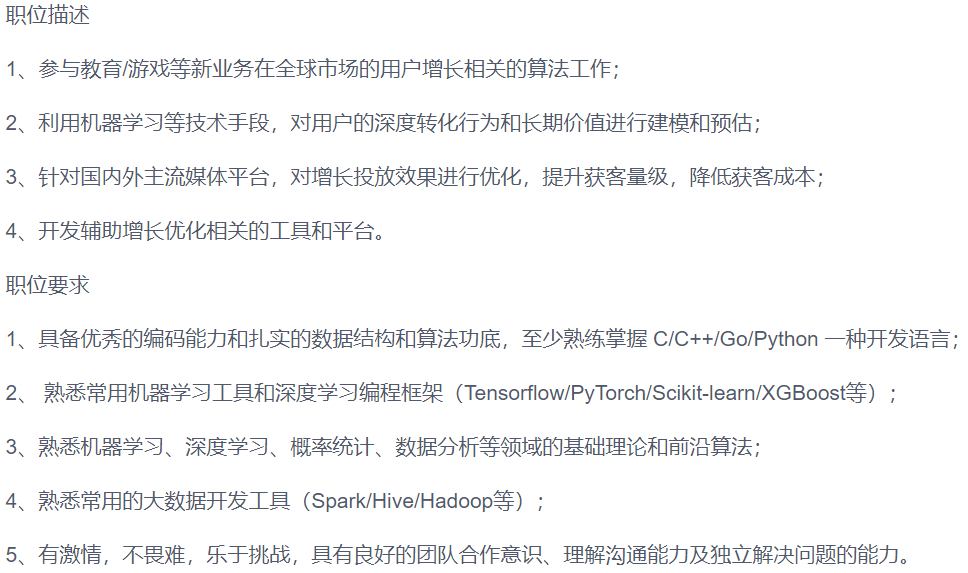
**第4章 专业知识在行业中的实践路径**

下面以两个例子：百度和字节跳动的机器学习工程师招聘信息，来进行分析。

以下为百度的招聘信息：



以下为字节跳动的招聘信息：



可以看出，机器学习工程师的职能就是对某一特定应用领域进行建模，并利用此模型进行数据的挖掘、预测、评估等，例如在搜索场景中的应用和对用户的评估。

**第5章 就业准备建议**

## 5.1 认清机器学习的趋势——垂直应用正在兴起

这几年机器学习已经成为行业热点和学术热点，有人说2019年或者2020年是数据元年，可以看出其火热程度。那么这个领域有什么趋势呢？其中之一就是垂直应用正在兴起。

之前，机器学习一直在算法领域发展。近一两年，不同的垂直领域开始更加深度的应用机器学习。那么未来，机器学习会实实在在的应用到各个行业中，比如金融、营销、销售、医疗，现在很多公司在做这种垂直领域内的应用，也有物流、金融专业背景的人进入这个领域。

## 5.2 认识机器学习领域的岗位——不仅仅是算法工程师

其实算法工程师仅仅这个领域的一小部分，相关岗位门类很多。整个机器学习的项目流程中，会涉及到原始的数据采集、如何理解、形式化业务问题，如何构建机器学习数据库，如何找模型，如何给模型调参，如何部署模型，以及部署之后的性能检测、在线评估、持续学习等等。所以整个机器学习的生命周期非常长，会包括很多门类的技术岗位。

## 5.3 求职机器学习的准备——从语言到心态

想要应聘机器学习研发或工程师类的职位，首先要通过技术要求，比如编程语言、对算法理解、对模型的理解，对目前流行的开发框架的理解。特别是对算法背后理论知识的理解，能够看出你解决真实问题的能力。

所以招聘中，项目经历或比赛经历是这类岗位看重的基础应用能力，然后就是理论知识，比如为什么有一个性能很好的模型应用到实际项目中却会变得性能很差，你能不能理解这个模型对理想数据分布的假设是什么？和真实生产数据会有哪些不同。所以有扎实的理论基础，就很重要。这里也再次强调一下数学知识，比如线性代数、微积分、概率论与数理统计等都是必备的基础学科知识。很多机器学习问题，都可以转化为某种约束条件下的优化问题，所以对数学的学习是绕不开的。

而软技能方面，要对学习、探索新知识、新技术有极大的热情。IT领域中各种技术的快速演进和迭代，必然要求技术类人员持续、快速地学习，这是行业特点要求的。所以要首先问问自己，是不是喜欢学习不断地新技术，有没有对新知识的热情和渴望。

**结 论**

机器学习领域经过几年的发展，虽然势头没有那么猛，但是市场需求依旧很大、前景依旧很好，就业形势很乐观。只要扎实修炼自身，从基础、专业知识、应用等方面，提升个人能力，就不愁找到一份理想的工作。