ശാസ്ത്രം ▶

ഡോ. കെ.കെ. ജോസ്

ഡേറ്റാ സയൻസും
ഡേറ്റാ അനലിറ്റിക്സും

ഇന്റർനെറ്റിന്റെയും മൊബൈൽ ഫോണിന്റെയും സോഷ്യൽ മീഡിയയുടെയുമൊക്കെ അഭൂതപൂർവ്വമായ വളർച്ചമൂലം ഡാറ്റാ അഥവാ വിവരങ്ങൾ/വസ്തുതകൾ നമ്മുടെ മസ്തിഷ്കത്തിൽ കുമിഞ്ഞു കൂടുകയാണ്. വിവിധ സ്രോതസ്സുകളിൽനിന്ന് വിവരശേഖരണം നടത്തി, അതിനെ വിശകലനം ചെയ്ത് പുതിയ പ്രവണതകളും ക്രമങ്ങളും കണ്ടെത്തി അവതരിപ്പിക്കുകയും അതിനെ അടിസ്ഥാനമാക്കി തീരുമാനങ്ങളിലെത്തിച്ചേരുകയും ചെയ്യുന്ന പ്രക്രിയയായ ഡേറ്റാ സയൻസിനെ പരിചയപ്പെടുത്തുന്ന ലേഖനം.

ഇന്നത്തെ കാലഘട്ടത്തെ ഡേറ്റായുഗം എന്നു പറയാം. ഇന്ന് എല്ലാം നിയന്ത്രിക്കുന്നത് ഡേറ്റയാണ്. ഡേറ്റാ എന്നാൽ വിവരം (information) നല്കുന്ന എന്തുമാണ്. ഇപ്പോൾ ബിഗ് ഡേറ്റായുടെ കാലമാണ്. നമുക്കു ലഭിക്കുന്ന സന്ദേശങ്ങളും ഇ-മെയിലുകളും ഫെയ്സ്ബുക്ക്, ടിറ്റർ, വാട്സ്ആപ്പ് തുടങ്ങിയ സോഷ്യൽ മീഡിയ വഴി ലഭിക്കുന്ന എന്തും, ഇന്റർനെറ്റ് ചാറ്റിംഗ്, സെർച് എൻജിനുകളും സെർവറുകളും വെബ്സൈറ്റുകളും ഇന്റർനെറ്റ് ഓഫ് തിംഗ്സ് വഴി ഉത്പാദിപ്പിക്കുന്ന വിവരങ്ങളും ഫോട്ടോകളും എസ്.എം.എസ്കളും ടെക്സ്റ്റ് ഡോക്യുമെന്റുകളും എല്ലാം ബിഗ്ഡേറ്റായുടെ ഭാഗമാണ്. ഉദാഹരണത്തിന് നമ്മുടെ കൈവശമുള്ള മൊബൈൽഫോണും ലാപ്ടോപ്പും കമ്പ്യൂട്ടറും ബിഗ്ഡേറ്റകളുടെ ശേഖരമാണ്. 2020-ൽ ദിവസംതോറും 44 സീറ്റാബൈറ്റ്സ്(Zettabytes, 1 zetta byte=2^{70}) ഡേറ്റാ ഉത്പാദിപ്പിച്ചിരുന്നെങ്കിൽ 2025 ആകുമ്പോൾ ദിനംപ്രതി 463 എക്സാബൈറ്റ്സ്(Exabytes) വീതം ഡേറ്റാ ഉത്പാദിപ്പിക്കപ്പെടും എന്നാണ് കണക്കുകൂട്ടലുകൾ പറയുന്നത്.

ഇത്തരം ബിഗ്ഡേറ്റാകളുടെ പ്രവാഹത്തെ പഠിക്കുന്നതിനും അവയുടെ രീതികളും പാറ്റേണുകളും മനസ്സിലാക്കി അപഗ്രഥിക്കുന്നതിനും പ്രവചിക്കുന്നതിനും തീരുമാനങ്ങൾ എടുക്കുന്നതിനുമായി അടുത്തകാലത്ത് ആവിർഭവിച്ചിരിക്കുന്ന വിജ്ഞാനശാഖകളാണ് ഡേറ്റാസയൻസും ഡേറ്റാ അനലിറ്റിക്സും മറ്റും. ഓരോ ദിവസവും ഡിജിറ്റൽ ലോകത്ത് ഉത്പാദിപ്പിക്കപ്പെടുന്ന ബിഗ്ഡേറ്റായുടെ അപഗ്രഥനത്തിനും കൈകാര്യം ചെയ്യലിനും പരമ്പരാഗത സ്റ്റാറ്റിസ്റ്റിക്കൽ ഡേറ്റാ അപഗ്രഥനരീതികൾ പര്യാപ്തമല്ല. ഇവിടെയാണ് കമ്പ്യൂട്ടർ പ്രോഗ്രാമിംഗ്, അൽഗോരിതങ്ങൾ ഡേറ്റാ വെയർഹൗസിംഗ്, ഡേറ്റാ മൈനിംഗ്, ആർട്ടിഫിഷ്യൽ ഇന്റലിജൻസ്, മെഷീൻലേണിംഗ് തുടങ്ങിയ ആധുനിക ശാസ്ത്രശാഖകളുടെ പ്രസക്തി. ഗണിതശാസ്ത്രം, സ്റ്റാറ്റിസ്റ്റിക്സ്, കമ്പ്യൂട്ടർ സയൻസ്, ഇൻഫർമേഷൻ സയൻസ്, ആർട്ടിഫിഷ്യൽ ഇന്റലിജൻസ്, മെഷീൻ ലേണിംഗ്, ഡേറ്റാ മൈനിംഗ് എന്നിവയുടെ സമ്മിശ്രശാഖയാണ് ഡേറ്റാ സയൻസ്.

ഡേറ്റാ അനലിറ്റിക്സ്, സ്റ്റാറ്റിസ്റ്റിക്സ്, മാത്തമാറ്റിക്സ്, ഡേറ്റാ അനലിസിസ് എന്നിവ വഴി കൃത്യമായ നിഗമനങ്ങളും തീരുമാനങ്ങളും എടുക്കുവാൻ സഹായിക്കുന്നു. ഡേറ്റാ സയൻസിന്റെ ഒരു ഉപശാഖയാണ് ഡേറ്റാ അനലിറ്റിക്സ് എന്നുപറയാം. ഡേറ്റാ സയൻസിൽ പാറ്റേണുകളും ബന്ധങ്ങളും രീതികളും മനസ്സിലാക്കി കൂടുതൽ അപഗ്രഥനത്തിനായി തരംതിരിക്കപ്പെടുന്ന ഡേറ്റയെ സ്റ്റാറ്റിസ്റ്റിക്കൽ അപഗ്രഥനരീതിവഴി വ്യക്തമായ നിഗമനങ്ങളിലേക്കും തീരുമാനങ്ങളിലേക്കും നയിക്കുന്നതിന് ഡേറ്റാ അനലിറ്റിക്സ് സഹായിക്കുന്നു. ബൃഹത്തായ ഡേറ്റാകളുടെ അപഗ്രഥനം വേഗത്തിൽ നടത്തുന്നതിന് കമ്പ്യൂട്ടർ പ്രോഗ്രാമുകളും അൽഗോരിതങ്ങളും സഹായിക്കുന്നു. ഇതിനായി പൈതൺ, ജൂലിയ, ആർ പ്രോഗ്രാം മുതലായ സ്വതന്ത്ര പ്രോഗ്രാമിംഗ് ഭാഷകളും ഉപയോഗിക്കുന്നു. അതായത് ഡേറ്റാകളിൽ നിന്നും ആഴത്തിൽ പഠിച്ച് വേഗത്തിൽ തീരുമാനങ്ങളെടുക്കുന്ന ശാസ്ത്രമാണ് ഡേറ്റാസയൻസ്.

ഡേവിഡ് ഡോനോഹോ

ബിഗ്ഡേറ്റാകളുടെ പ്രവാഹത്തെ പഠിക്കുന്നതിനും അവയുടെ രീതികളും പാറ്റേണുകളും മനസ്സിലാക്കി അപഗ്രഥിക്കുന്നതിനും പ്രവചിക്കുന്നതിനും തീരുമാനങ്ങൾ എടുക്കുന്നതിനുമായി അടുത്തകാലത്ത് ആവിർഭവിച്ചിരിക്കുന്ന വിജ്ഞാനശാഖകളാണ് ഡേറ്റാസയൻസും ഡേറ്റാ അനലിറ്റിക്സും

ഡേറ്റാസയൻസ് വഴി ബുദ്ധിപരമായ ചോദ്യങ്ങൾ രൂപീകരിക്കുകയും ഡേറ്റാ അനലിറ്റിക്സ് വഴി ഡേറ്റാ അപഗ്രഥനം നടത്തി നിഗമനങ്ങളിൽ എത്തുകയും ചെയ്യുന്നു. പലരും ഡേറ്റാ സയൻസും ഡേറ്റാ അനലിറ്റിക്സും ഒന്നുതന്നെയാണ് എന്നു കരുതുന്നു, കാരണം രണ്ടും ബിഗ്ഡേറ്റായാണ് കൈകാര്യം ചെയ്യുന്നത്. പക്ഷേ, ഒരു നാണയത്തിന്റെ രണ്ട് വശങ്ങളാണ് ഇവ എന്നു പറയാം. ഡേറ്റാ സയൻസ് ബിഗ്ഡേറ്റയിൽനിന്ന് അതിവേഗത്തിൽ അർത്ഥവത്തായ ചില സൂചനകളും ഉത്തരം കിട്ടേണ്ട ചോദ്യങ്ങളും രൂപീകരിച്ചു നല്കുന്നു. ഡേറ്റാ അനലിറ്റിക്സ് അവയ്ക്ക് സ്റ്റാറ്റിസ്റ്റിക്സും മാത്തമാറ്റിക്സും ഡേറ്റാ അനാലിസിസ് സോഫ്ട്വെയറുകളും ഉപയോഗിച്ച് നിർദ്ധാരണം ചെയ്ത് ഉത്തരങ്ങളും നിഗമനങ്ങളും നൽകി ബുദ്ധിപൂർവകമായ തീരുമാനങ്ങളിലെത്താൻ സഹായിക്കുന്നു.

ഡേവിഡ് ഡോനോഹോ എന്ന പ്രശസ്ത ഡേറ്റാ സയന്റിസ്റ്റിന്റെ 2015-ൽ പ്രസിദ്ധീകരിച്ച 'ഡേറ്റാ സയൻസിന്റെ

അമ്പത് വർഷങ്ങൾ' എന്ന ലേഖനത്തിൽ ഇതിന്റെ ആവിർഭാവത്തെപ്പറ്റിയും ആവശ്യകതയെപ്പറ്റിയും സ്റ്റാറ്റിസ്റ്റിക്സും ഡേറ്റാ സയൻസും തമ്മിലുള്ള സാമ്യവും വ്യത്യാസങ്ങളും സംബന്ധിച്ചും വിശദമായി പ്രതിപാദിക്കുന്നുണ്ട്. 1962-ലാണ് ജോൺടക്കി എന്ന പ്രശസ്തനായ സ്റ്റാറ്റിസ്റ്റിഷ്യൻ, ഡേറ്റാ അനലിസിസിൽ വരാവുന്ന വെല്ലുവിളികളെപ്പറ്റി മുന്നറിയിപ്പ് തന്നതും അതിനായി സജ്ജരാകുവാൻ ലോകമെങ്ങുമുള്ള സ്റ്റാറ്റിസ്റ്റിക്സ് വിദഗ്ധരെ ഉപദേശിച്ചതും. ടക്കിയുടെ മുന്നറിയിപ്പിന്റെ 60-ാം വാർഷികത്തിൽ ഈ ലേഖനം വളരെ പ്രസക്തമാണ്.

ജോൺ ചേമ്പേഴ്സ്

2000 വരെ സ്റ്റാറ്റിസ്റ്റിക്സ് പണ്ഡിതർ ഇതിന് വേണ്ടത്ര പ്രാധാന്യം നൽകിയില്ല എന്നതാണ് യാഥാർത്ഥ്യം. 2001-ൽ ബ്രയ്മാൻ സ്റ്റാറ്റിസ്റ്റിക്കൽ സയൻസ് എന്ന ജേർണലിൽ ഇക്കാര്യം വ്യക്തമാക്കുന്നുണ്ട്. തിയറികളിൽ അധിഷ്ഠിതമായ ജനറേറ്റീവ് മോഡലിംഗിൽ വിശ്വസിച്ചവരും പ്രഡിക്ടീവ് മോഡലിംഗിൽ വിശ്വസിച്ച് ഡേറ്റാ നിയന്ത്രിത മോഡലുകൾക്ക് രൂപം കൊടുത്തവരും തമ്മിൽ വലിയ വിടവുണ്ടായതായി അദ്ദേഹം പറയുന്നു. രണ്ടു സംസ്കാരങ്ങൾ പോലെ ഇവ തുടരുകയും 2009-ൽ മാർക്ക് ലിബർമാൻ ഒരു കോമൺ ടാസ്ക് ഫ്രെയിംവർക്കിന് രൂപംകൊടുക്കുകയും ചെയ്തു. ഇതിന്റെ ഫലമായി കമ്പ്യൂട്ടർ പ്രോഗ്രാമുകൾ ഉപയോഗിച്ച് മെഷീൻ ലേണിംഗ് സാങ്കേതികവിദ്യ രൂപം കൊണ്ടു. ഇക്കാലത്തുതന്നെയാണ് പൈതൺ പ്രോഗ്രാമും R - പ്രോഗ്രാമും രൂപംകൊണ്ടതും പ്രചാരത്തിൽ വന്നതും. ഇവ സ്വതന്ത്ര സോഫ്റ്റ്‌വയറുകൾ ആയിരുന്നതുകൊണ്ട് വേഗം പ്രചരിക്കുകയും കൂടുതൽ ഉപയോഗപ്രദമായ മാറ്റങ്ങൾ ഉണ്ടാവുകയും ചെയ്തു. കമ്പ്യൂട്ടർ സയൻസ് ഗ്രൂപ്പാണ് ഇതിനു നേതൃത്വം കൊടുത്തത്. ഇത് പ്രഡിക്ടീവ് മോഡലിംഗ് രീതിയിൽ അധിഷ്ഠിതമായ ഡേറ്റാ സയൻസിന്റെ വളർച്ചയ്ക്ക് ആക്കം കൂട്ടി.

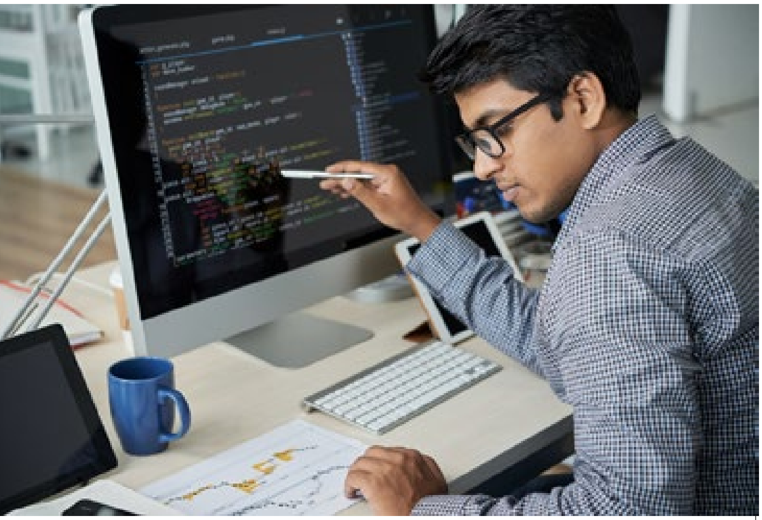
വില്യം ക്ലാവ്ലാൻഡ്

ഇതിന്റെ ഫലമായി 2015 സെപ്റ്റംബർ 11-ന് അമേരിക്കയിലെ യൂണിവേഴ്സിറ്റി ഓഫ് കാലിഫോർണിയ ബെർക്ക്ലി ക്യാംപസ്സിൽ ഡേറ്റാ സയൻസിൽ മാസ്റ്റേഴ്സ് പ്രോഗ്രാമിന് തുടക്കം കുറിച്ചു. താമസിയാതെതന്നെ ചേമ്പേഴ്സ്, ക്ലാവ്ലാൻഡ് എന്നീ ശാസ്ത്രജ്ഞർ 'ഗ്രേറ്റർ ഡേറ്റാ സയൻസ്' പ്രോജക്ടിന് രൂപംകൊടുത്തു. ആറ് വിവിധ ഡിവിഷനുകളുള്ള ബ്രഹത്തായ ഒരു പദ്ധതിയായിരുന്നു അത്. ഇതിൽ മൂന്ന് ഡിവിഷനുകൾ കംപ്യൂട്ടർ സയൻസ്സുമായും മൂന്ന് എണ്ണം സ്റ്റാറ്റിസ്റ്റിക്സുമായും ബന്ധപ്പെട്ടും ആയിരുന്നു. ബിഗ് ഡേറ്റാകളെ അതിവേഗം പ്രോസസ് ചെയ്യുന്നതിനും അപഗ്രഥനം ചെയ്യുന്നതിനും കമ്പ്യൂട്ടർ പ്രോഗ്രാമുകളും അൽഗോരിതങ്ങളും വിഷലൈസേഷൻ ടെക്നിക്കുകളും ആവശ്യമാണ്. ഇതാണ് ഒന്നാം ഘട്ടം എന്നു പറയാം. ഇത് കമ്പ്യൂട്ടർ സയൻസിന്റെ മേഖലയാണ്. രണ്ടാമത്തെ ഘട്ടത്തിൽ സ്റ്റാറ്റിസ്റ്റിക്സിലെ വിവിധ ടെക്നിക്കുകൾ ഉപയോഗിച്ച് ഡേറ്റാ അപഗ്രഥനം ചെയ്യുകയും മോഡലുകളും ടെസ്റ്റുകളും നടത്തുകയും ഭാവിയിലേക്കുള്ള ട്രെൻഡുകളും മറ്റും പ്രവചിക്കുകയും കണ്ടുപിടിക്കുകയും നിഗമനങ്ങളിലെത്തുകയും വേണം. ഇതാണ് ഡേറ്റാ അനലിറ്റിക്സ്. ഇത് പൊതുവേ സ്റ്റാറ്റിസ്റ്റിക്സ് വിദഗ്ധർ ചെയ്യേണ്ടതാണ്. രണ്ടും പ്രധാനമാണ്. അതുകൊണ്ട് ഡേറ്റാ സയൻസും ഡേറ്റാ അനലിറ്റിക്സും പരസ്പരം പൂരകങ്ങളായ രണ്ടു ശാസ്ത്രശാഖകളാണ്. തൊഴിൽ അവസരങ്ങൾ ഒട്ടനവധിയാണ്.

ഒരു ഡേറ്റാ സയന്റിസ്റ്റ് പ്രധാനമായും താഴെപ്പറയുന്ന കാര്യങ്ങൾക്കു നേതൃത്വം കൊടുക്കുന്നു: (1) ലഭിച്ചിരിക്കുന്ന ഡേറ്റാകളെ പരിശോധിച്ച് പാകപ്പെടുത്തി അപാകതകൾ നീക്കി ക്ലീൻ ചെയ്ത് ഗുണമേന്മയും കൃത്യതയും വേഗത്തിൽ ഉറപ്പാക്കുന്നു. (2) അന്വേഷണാത്മകവും ഗവേഷണപരവുമായ രീതിയിൽ ബിഗ് ഡേറ്റാകളെ പഠിച്ച് അവലോകനം ചെയ്യുന്നു. (3) ഇ.ടി.എൽ ശൃംഖലവഴി ഡേറ്റാ മൈനിംഗ് നടത്തുന്നു (4) ഗ്രാഫിക്സ് വഴി ദൃശ്യാവിഷ്കരണം നടത്തി പാറ്റേണുകൾ, കൂട്ടങ്ങൾ മുതലായവ കണ്ടെത്തുന്നു. (5) ഏതു വിവരവും കോഡുകൾ വഴി ഓട്ടോമോഷനു പര്യാപ്തമാക്കുന്നു. (6) റാൻഡം ഫോറസ്റ്റ്, ഡിസിഷൻ ട്രീസ് മുതലായവ വഴിയും മെഷീൻ ലേണിംഗ് അൽഗോരിതങ്ങൾ വഴിയും ഡേറ്റായിൽ നിന്നും നേരിട്ടുതന്നെ പുതിയ വിവരങ്ങൾ ശേഖരിക്കുന്നു. അതിനായി ഒരു ഡേറ്റാ സയന്റിസ്റ്റ് മാത്തമാറ്റിക്സ്, പ്രോബബിലിറ്റി, സ്റ്റാറ്റിസ്റ്റിക്സ്, ഗ്രാഫിക്സ്, പ്രോഗ്രാമിംഗ്, ഡേറ്റാ ബേസ് മാനേജ്മെന്റ്, ബിഗ് ഡേറ്റാ പ്ലാറ്റ്ഫോമുകളായ അപ്പാച്ചേ സ്പാർക്, ഹഡൂപ്പ് തുടങ്ങിയവ സംബന്ധിച്ച നല്ല അവഗാഹമുണ്ടാക്കണം.

ഒരു ഡേറ്റാ അനലിസ്റ്റ് ഡേറ്റാ ശേഖരണത്തിലും അപഗ്രഥനത്തിലും വിശദീകരണത്തിലും പ്രഗത്ഭനാകണം. എസ്.ക്യു.എൽ., ടാബ്ലോ തുടങ്ങിയവ വഴിയും എക്സൽ, സാസ്, പൈതൺ, ആർ, എസ്.പി.എസ്.എസ്. തുടങ്ങിയ സോഫ്ട്‌വെയറുകൾ ഉപയോഗിച്ച് ഡേറ്റാ അപഗ്രഥനം നടത്തുന്നതിലും സ്റ്റാറ്റിസ്റ്റിക്കൽ സാങ്കേതികവിദ്യകൾ വഴി നിഗമനങ്ങളിലെത്തുന്നതിനും സമർത്ഥനായിരിക്കണം. അതിനാൽ ഡേറ്റാ അനലിറ്റിക്സ് സ്റ്റാറ്റിസ്റ്റിക്സ് മേഖലയാണ്.

ഉദാഹരണത്തിന് കോവിഡ് സംബന്ധിച്ച ഡേറ്റാകൾ വിവിധ മാർഗങ്ങൾ ഉപയോഗിച്ച് ശേഖരിക്കുന്നതിനും പ്രായം, ലിംഗം, ഭക്ഷണരീതി, ജോലി, ജീവിതരീതി, മതം, സമ്പർക്കം, രാജ്യം, ജാതി, ആരോഗ്യസ്ഥിതി, മറ്റ് അസുഖ



ങ്ങൾ, കുടുംബപശ്ചാത്തലം തുടങ്ങി നിരവധി കാര്യങ്ങൾ അനുസരിച്ചുള്ള വകഭേദങ്ങൾ, ചികിത്സാരീതികൾ, പ്രത്യേകതകൾ തുടങ്ങിയവയെപ്പറ്റി അന്വേഷണാത്മകമായ രീതിയിൽ കമ്പ്യൂട്ടർ പ്രോഗ്രാമുകളും അൽഗോരിതങ്ങളും വഴി അതിവേഗം പാറ്റേണുകൾ, ട്രെൻഡുകൾ, ക്ലസ്റ്ററുകൾ, ഗ്രാഫുകൾ, ചാർട്ടുകൾ, ദൃശ്യമാർഗങ്ങൾ മുഖേന അർത്ഥവത്തായ കാര്യങ്ങളും ചോദ്യങ്ങളും വിശദപഠനത്തിനായി രൂപീകരിക്കുവാൻ സമർത്ഥനാകണം.

ഡേറ്റയിലെ തെറ്റുകളും പൊരുത്തക്കേടുകളും അവിശ്വസനീയതകളും ഗ്യാപ്പുകളും മാറ്റി എഡിറ്റുചെയ്ത് അപഗ്രഥനത്തിന് യോഗ്യമാക്കുന്നത് ഡേറ്റാ സയന്റിസ്റ്റാണ്. ചിലപ്പോൾ ചില കോഡുകളിലേക്കു മാറ്റുകയോ ഒക്കെ ചെയ്താലേ ഡേറ്റാ അപഗ്രഥനം സാധ്യമാകൂ. ഇതിന് ഡേറ്റാ ക്ലീനിംഗ്, ഡേറ്റാ ചികിത്സയും മാനേജ്മെന്റും എന്നൊക്കെ പറയുന്നു.

ബിസിനസ് അനലിറ്റിക്സ് ബിസിനസ് രംഗത്തെ ഗതിവിഗതികൾ സംബന്ധിച്ച് ആഴത്തിൽ മനസ്സിലാക്കുന്നു. കസ്റ്റമേഴ്സിന്റെ പ്രായം, ലിംഗം, തൊഴിൽ, വരുമാനം, ഭക്ഷണരീതി, വാസസ്ഥലം, സ്വത്ത്, വിദ്യാഭ്യാസം, ജീവിതനിലവാരം തുടങ്ങിയവയ്ക്കനുസരിച്ച് ആവശ്യങ്ങളിൽ വരുന്ന വ്യതിയാനങ്ങളെപ്പറ്റിയും ഉപഭോക്താക്കളുടെ താത്പര്യങ്ങളെക്കുറിച്ചും പഠിച്ച് നിഗമനങ്ങളിലെത്തി ബുദ്ധിപൂർവകമായ തീരുമാനങ്ങൾ എടുക്കണം. മെഡിക്കൽ അനലിറ്റിക്സിൽ രോഗികളെപ്പറ്റിയും രോഗത്തെപ്പറ്റിയും അവരുടെ പ്രായം, ലിംഗം, ജീവിതചര്യ, ഭക്ഷണം, വിദ്യാഭ്യാസ നിലവാരം, തൊഴിൽ, അസുഖങ്ങൾ, ആരോഗ്യനിലവാരം തുടങ്ങിയവ സംബന്ധിച്ച വിവിധങ്ങളായ വിവരങ്ങൾ ശേഖരിച്ച് യാഥാർത്ഥ്യബോധമുള്ള തീരുമാനങ്ങളെടുക്കുവാനും നടപ്പിലാക്കുവാനും ആരോഗ്യനിലവാരം ഉറപ്പാക്കുവാനും ഇത് സഹായിക്കുന്നു.

കാലാവസ്ഥാ നിരീക്ഷണത്തിലും പ്രവചനത്തിലും ആദായനികുതി വിവരശേഖരണത്തിനും സാമ്പത്തികരംഗത്തെ നിരവധി വിവരങ്ങൾ ശേഖരിച്ച് നടപടികൾ സ്വീകരിക്കുന്നതിനും ജനാഭിപ്രായം മനസ്സിലാക്കി സർക്കാർ നയങ്ങൾ രൂപവത്കരിക്കുന്നതിനും രാഷ്ട്രീയപാർട്ടികൾ ജനങ്ങളുടെ താല്പര്യം മനസ്സിലാക്കി പ്രചാരണം നടത്തുന്നതിനും ഡേറ്റാ സയൻസും ഡേറ്റാ അനലിറ്റിക്സും വ്യാപകമായി ഉപയോഗിക്കുന്നു. വ്യവസായരംഗത്ത് ഗുണനിലവാരം ഉറപ്പാക്കുന്നതിനും ഉല്പന്നങ്ങളെ മാർക്കറ്റ് ചെയ്യുന്നതിനും സ്റ്റോർ ചെയ്യുന്നതിനും ജനാഭിരുചി മനസ്സിലാക്കി ആകർഷകമാക്കി ഓരോ പ്രായക്കാർക്കും ലഭ്യമാക്കുന്നതിനും ലാഭം കൊയ്യുന്നതിനും ഡേറ്റാ സയൻസും ഡേറ്റാ അനലിറ്റിക്സും ഉപയോഗിക്കാം. ജനങ്ങളുടെ തൊഴിലില്ലായ്മ സംബന്ധിച്ചും കാർഷികരംഗത്തെ പ്രശ്നങ്ങൾ സംബന്ധിച്ചും വ്യക്തമായ വിവരങ്ങൾ ശേഖരിച്ച് പരിഹാര മാർഗങ്ങൾ ആസൂത്രണം ചെയ്യുന്നതിനും ഇവ സഹായിക്കും.

കോവിഡ് കാലത്ത് ഓൺലൈൻ വിദ്യാഭ്യാസം സാധ്യമാക്കിയതും ഡേറ്റാ സയൻസ്, ഡേറ്റാ അനലിറ്റിക്സ് മേഖലകളിലെ മുന്നേറ്റമാണ്. ഓൺലൈൻ - ഓഫ്‌ലൈൻ വിദ്യാഭ്യാസ രീതികളുടെ ഗുണങ്ങളും പോരായ്മകളും മനസ്സിലാക്കുന്നതിനും അധ്യാപകരുടെയും മാതാപിതാക്കളുടെയും വിദ്യാർത്ഥികളുടെയും അഭിപ്രായങ്ങൾ മനസ്സിലാക്കി കൂടുതൽ ഫലവത്തായ വിദ്യാഭ്യാസ രീതികളും മൂല്യനിർണയമാർഗങ്ങളും രൂപവത്കരിക്കുന്നതിനും ഇവ സഹായകമാണ്.

ട്രെയിൻ - വിമാന യാത്രകൾക്ക് ബുക്കിംഗിനും ക്യാൻസലേഷനും ലഗേജ് വിതരണത്തിനും തുടർ യാത്രാ ബുക്കിംഗിനും മറ്റും ട്രാൻസ്പോർട്ടേഷൻ കമ്പനികൾ ആശ്രയിക്കുന്നത് ഈ നൂതനസാങ്കേതികവിദ്യയാണ്. രാജ്യസുരക്ഷയ്ക്കും ഗവേഷണത്തിനും ഡേറ്റാ സയൻസ് മേഖല വിവിധ തലങ്ങളിൽ ഉപയോഗപ്രദമാണ്. ഭക്ഷ്യധാന്യങ്ങളുടെ ശേഖരണത്തിനും വിതരണത്തിനും കാര്യക്ഷമമായ പബ്ലിക് ഡിസ്ട്രിബ്യൂഷൻ ഉറപ്പുവരുത്തുന്നതിനും ഇവ സഹായിക്കും. ബഹിരാകാശ ഗവേഷണത്തിൽ ഡേറ്റാ സയൻസും ഡേറ്റാ അനലിറ്റിക്സും വൻ സാധ്യതകൾക്കു പ്രതീക്ഷ നല്കുന്നു. ബാങ്കിംഗ് - ഫിനാൻസ് - സ്റ്റോക്ക് മാർക്കറ്റിംഗ് മേഖലകളിൽ അനലിറ്റിക്സ് സാങ്കേതികവിദ്യ കൂടുതൽ ഫലവത്തായി ഉപയോഗിക്കാം. സൈബർ കുറ്റങ്ങൾ, ഡിജിറ്റൽ കറൻസി ഉപയോഗത്തിലുള്ള ഭീഷണികൾ മുതലായവ സംബന്ധിച്ച അന്വേഷണത്തിനും പഠനത്തിനും ഗവേഷണത്തിനും ഇവ സഹായകമാണ്.

രാജ്യഭരണത്തിനും രാഷ്ട്രതന്ത്രജ്ഞതയ്ക്കും അനലിറ്റിക്സ് വളരെ സഹായകമാണ്. വിവിധ വകുപ്പുകളെ ഒരുമിച്ച് പ്രവർത്തിപ്പിക്കുന്നതിനും പരസ്പരം കോർത്തിണക്കി നയരൂപീകരണം നടത്തി നടപ്പിലാക്കുന്നതിനും ഫലപ്രാപ്തിയെപ്പറ്റി തുടർപഠനം നടത്തുന്നതിനും ഡേറ്റാ സയൻസ് ഉപയോഗപ്രദമാണ്. മെഡിക്കൽ രംഗത്തും ഫാർമസ്യൂട്ടിക്കൽ രംഗത്തും ചികിത്സയ്ക്കും ക്ലിനിക്കൽ ട്രയൽസ് നടത്തുന്നതിനും മെറ്റാ അനാലിസിസ് വഴി വിവിധ പഠനങ്ങളെ സംയോജിപ്പിച്ച് കൂടുതൽ കൃത്യമായ നിഗമനങ്ങളിലെത്തുന്നതിനും ഇത് സഹായിക്കും.

ലേഖകൻ: ഡയറക്ടർ, സ്കൂൾ ഓഫ് മാത്തമാറ്റിക്സ് ആൻഡ് സ്റ്റാറ്റിസ്റ്റിക്സ്, സ്കൂൾ ഓഫ് ഡാറ്റാ അനലിറ്റിക്സ്,മഹാത്മാഗാന്ധി യൂണിവേഴ്സിറ്റി,കോട്ടയം