**Зертханалық жұмыс №1: Orange бағдарламасында деректер жиынын жүктеу, тазалау және деректерді алдын ала өңдеу**

**Мақсаты:**
Orange бағдарламасында деректер жиынын жүктеу, жетпейтін мәндерді өңдеу, деректерді масштабтау және нормализациялау сияқты алдын ала өңдеу қадамдарын орындауды үйрену.

**1. Қажетті құралдар:**

* Orange бағдарламалық жасақтамасы
* CSV немесе Excel форматындағы деректер жиыны (мысалы, Iris деректер жиыны)

**2. Зертханалық жұмыстың қадамдары:**

**2.1 Деректер жиынын жүктеу**

1. **Orange бағдарламасын ашыңыз.**
2. **File** виджетін қосыңыз:
	* Сол жақтағы құралдар панелінен **File** виджетін сүйреңіз және жұмыс аймағына орналастырыңыз.
3. **File** виджетін екі рет шертіп, қажетті деректер жиынын таңдаңыз:
	* Егер сізде дайын деректер болмаса, Orange-де кіріктірілген **Iris** немесе басқа деректер жиындарын қолдануға болады.
	* Деректердің форматтары: CSV, Excel немесе табуляция арқылы бөлінген деректер.

**2.2 Деректер жиынын қарау**

1. Деректер жүктелгеннен кейін, **Data Table** виджетін қосыңыз:
	* **File** виджетінен **Data Table** виджетіне сызық жүргізіп, деректерді визуалдаңыз.
2. **Data Table** виджетін ашып, деректерді қарап шығыңыз:
	* Деректердің жалпы құрылымын тексеріңіз (сандық және мәтіндік атрибуттар, жетпейтін мәндер).

**2.3 Деректерді тазалау**

1. **Select Columns** виджетін қосыңыз:
	* Бұл виджет арқылы қажетті немесе қажетсіз атрибуттарды алып тастауға болады.
	* Мысалы, егер деректерде зерттеуге қатысы жоқ атрибуттар болса, оларды алып тастай аласыз.
2. **Missing Data** виджетін қосыңыз:
	* Жетпейтін мәндер бар-жоғын тексеру үшін **Missing Data** виджетін қолданыңыз.
	* Егер жетпейтін мәндер анықталса, оларды әртүрлі жолмен өңдеуге болады: мәндерді медианамен ауыстыру, орташа мәнді қолдану, немесе сол жазбаларды алып тастау.
3. **Missing Data** виджетін ашыңыз және өңдеу әдісін таңдаңыз:
	* Мысалы, жетпейтін мәндерді бағанның орташа мәнімен алмастыруды таңдаңыз.

**2.4 Деректерді нормализациялау және масштабтау**

1. **Preprocess** виджетін қосыңыз:
	* Бұл виджет арқылы деректерді масштабтау немесе нормализациялауға болады.
2. **Preprocess** виджетін ашыңыз және төмендегілерді орындаңыз:
	* **Scaling**: Деректерді орташа мәнді 0, стандартты ауытқуды 1 болатындай етіп стандарттау.
	* **Normalization**: Деректерді [0, 1] диапазонына келтіру.

**2.5 Деректерді визуализациялау**

1. **Scatter Plot** немесе **Box Plot** виджетін қосыңыз:
	* Алдын ала өңделген деректерді визуалдап, нәтижелерді көріңіз.
2. **Scatter Plot** виджетін ашыңыз және әртүрлі атрибуттарды осьтерге орналастыру арқылы деректердің таралуын бақылаңыз.

**2.6 Алынған нәтижелерді сақтау**

1. **Save Data** виджетін қосыңыз:
	* Өңделген деректерді файл ретінде сақтаңыз.
	* CSV немесе Excel форматында сақтауға болады.

**3. Қорытынды сұрақтар:**

1. Жетпейтін мәндерді қандай әдістермен өңдедіңіз? Қайсысы тиімді болды деп ойлайсыз?
2. Деректерді нормализациялаудың артықшылықтары қандай?
3. Алдын ала өңделген деректердің классификация немесе регрессия сияқты кейінгі талдауларға қалай әсер етуі мүмкін?

**4. Зертханалық жұмыс бойынша есеп:**

* Деректерді алдын ала өңдеу қадамдарын сипаттаңыз.
* Жетпейтін мәндерді қалай өңдегеніңізді түсіндіріңіз.
* Деректерді масштабтау немесе нормализациялау нәтижелерін талдаңыз.
* Визуализация нәтижелерін қосып, түсініктеме беріңіз.