

ChemSuite

A complete package for Chemoinformatics
calculations and Machine Learning

Contents

1	Introduction.....	1
2	Installation step for window user.....	2
2.1	JAVA.....	2
2.1.1	How to copy path.....	2
2.1.2	Windows 7 path setting.....	3
2.1.3	Windows 10 path setting.....	4
2.2	Anaconda.....	5
2.3	ChemSuite-Installer.....	7
3	Installation steps for Linux user.....	13
4	ChemSuite Tutorial.....	14
4.1	Home Page.....	14
4.2	Preprocessing Panel.....	16
4.3	Calculation Tab.....	17
4.4	Processing Tab.....	19
4.5	Normalizaton.....	20
4.6	Feature selection.....	22
4.7	Development & Validation.....	23
4.8	Predict.....	25
4.9	Analysis.....	26

1 Introduction

At a basic level, machine learning is about predicting the future based on the past. For instance, you might wish to predict how much a user Alice will like a movie that she hasn't seen, based on her ratings of movies that she has seen. This means making informed guesses about some unobserved property of some object, based on observed properties of that object. The first question we'll ask is: what does it mean to learn? In order to develop learning machines, we must know what learning actually means, and how to determine success (or failure).

Ever since computers were invented, we have wondered whether they might be made to learn. If we could understand how to program them to learn-to improve automatically with experience-the impact would be dramatic. Imagine computers learning from medical records which treatments are most effective for new diseases, houses learning from experience to optimize energy costs based on the particular usage patterns of their occupants, or personal software assistants learning the evolving interests of their users in order to highlight especially relevant stories from the online morning newspaper. A successful understanding of how to make computers learn would open up many new uses of computers and new levels of competence and customization. And a detailed understanding of information processing algorithms for machine learning might lead to a better understanding of human learning abilities (and disabilities) as well.

We do not yet know how to make computers learn nearly as well as people learn. However, algorithms have been invented that are effective for certain types of learning tasks, and a theoretical understanding of learning is beginning to emerge. Many practical computer programs have been developed to exhibit useful types of learning, and significant commercial applications have begun to appear. For problems such as speech recognition, algorithms based on machine learning outperform all other approaches that have been attempted to date. In the field known as data mining, machine learning algorithms are being used routinely to discover valuable knowledge from large commercial databases containing equipment maintenance records, loan applications, financial transactions, medical records, and the like. As our understanding of computers continues to mature, it seems inevitable that machine learning will play an increasingly central role in computer science and computer technology.

Definition

Machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new – but one that's gaining fresh momentum.

Because of new computing technologies, machine learning today is not like the past. While many machine learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data – over and over, faster and faster – is a recent development. Here are a few widely publicized examples of machine learning applications that you may be familiar with:

- To solve science related problem like handling large amount of data.
- The heavily hyped, self-driving Google car? The essence of machine learning.
- Knowing what customers are saying about you on Twitter? Machine learning combined with linguistic rule creation.

2 Installation step for window user

To use ChemSuite software a user needs to install some pre-requisite packages listed below in Table 1.

Table 1 List of packages for ChemSuite

S/N	Package	Source
1.	Java (9.0.1)	http://www.oracle.com/technetwork/java/javase/downloads/index.html
2.	Anaconda (5.0.1)	https://www.anaconda.com/download/
3.	PyDPI	https://pypi.python.org/pypi/conda/3.17.0
4.	Microsoft visual C++ compiler	https://www.microsoft.com/en-in/download/details.aspx?id=44266
5.	RDkit	https://anaconda.org/rdkit/rdkit
6.	Pytz	https://anaconda.org/anaconda/pytz
7.	Sci-kit learn	https://anaconda.org/anaconda/scikit-learn
8.	ChemSuite (0.1)	http://14.139.57.41/chemsuiteweb/download.php

A user can download all packages manually or follow steps given below to complete the installation process.

Important notice: - Please first remove Python, if you have already installed into your system then start installation steps.

2.1 JAVA

The first step is to install Java, if you do not have in your computer follow simple steps as shown in Figure 1 to download Java, type “jdk” in google.

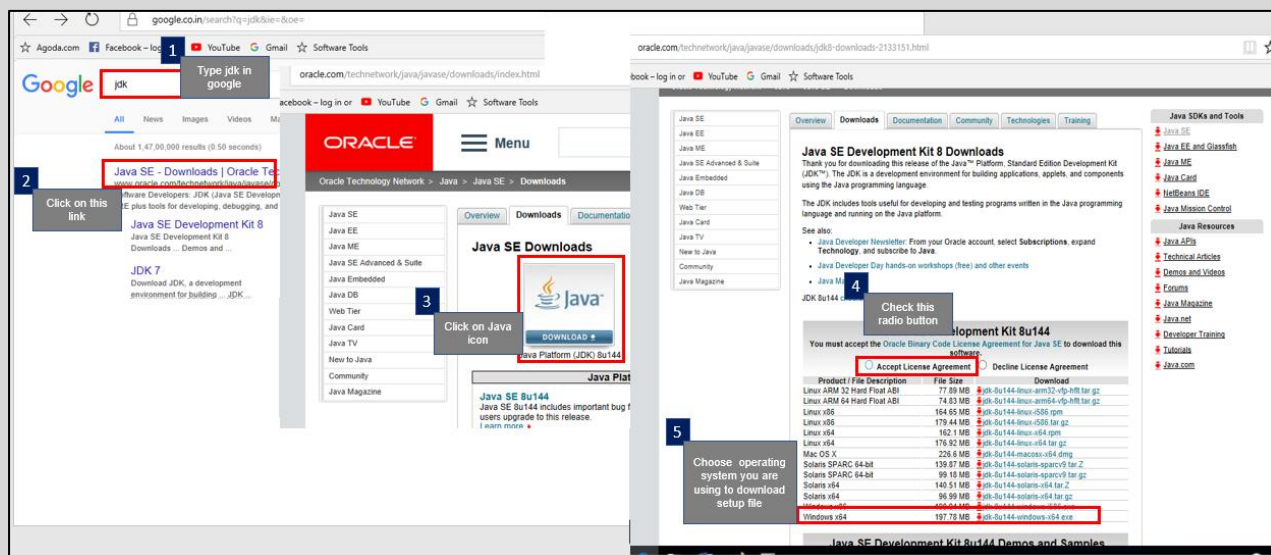


Figure 1

2.1.1 How to copy path

1. Double click on **My computer**.
2. Go to **C:/ drive** double click it.
3. Then go to **Program files**
4. Next go to **Java** folder
5. Double click on **jdk**
6. Then double click on **bin** folder

ChemSuite Tutorial

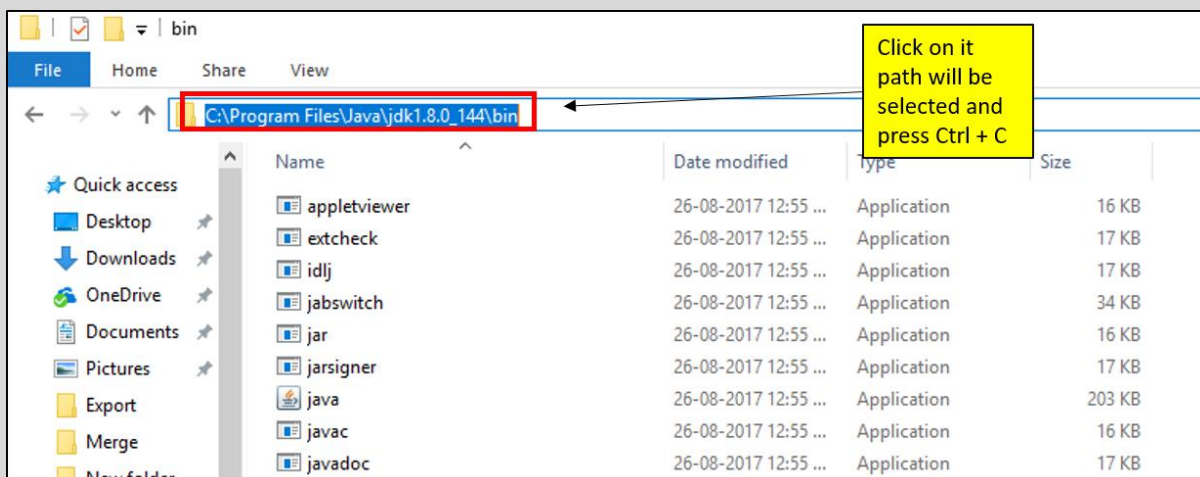


Figure 2

7. Repeat same process to copy **ire** path.

2.1.2 Windows 7 path setting

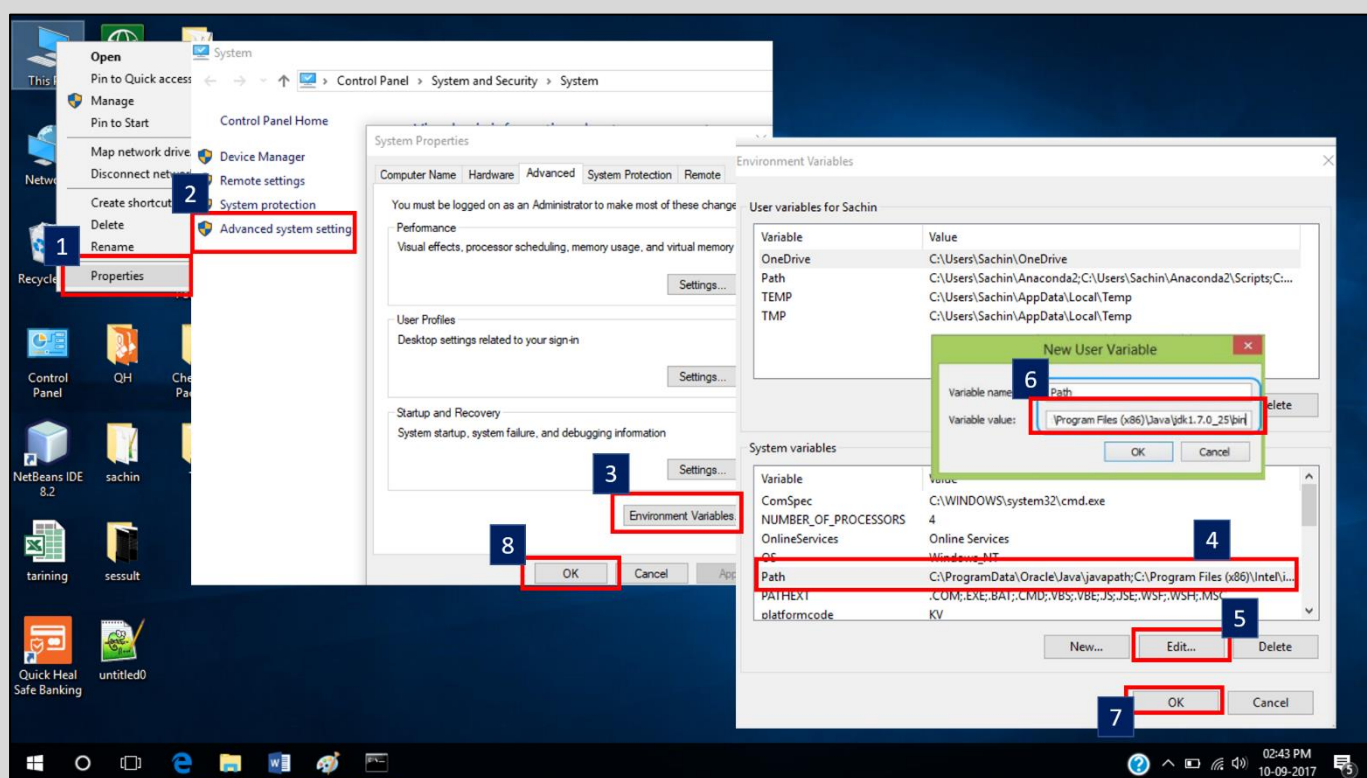


Figure 3

Follow steps as shown in Figure 3 to set path into environment variable.

1. **Properties**
2. **Advanced system settings**
3. **Environment variable**

ChemSuite Tutorial

- Click on **Path**
- Click on **Edit**
- For **windows 7** put a semicolon before path like shown in the example below in red color and then the path in yellow color paste whole red and yellow color path without space inside double quote into **step 6**.

“C:\Program Files\Java\jdk1.8.0_144\bin ; C:\Program Files\Java\jre1.8.0_144\bin”

“;C:\Program Files\Java\jdk1.8.0_144\bin;C:\Program Files\Java\jre1.8.0_144\bin”

2.1.3 Windows 10 path setting

In window 10 step to set path into environment variable are described in Figure 4.

- Properties**
- Advanced system settings**
- Environment variable**
- Click on **Path**
- Click on **Edit**

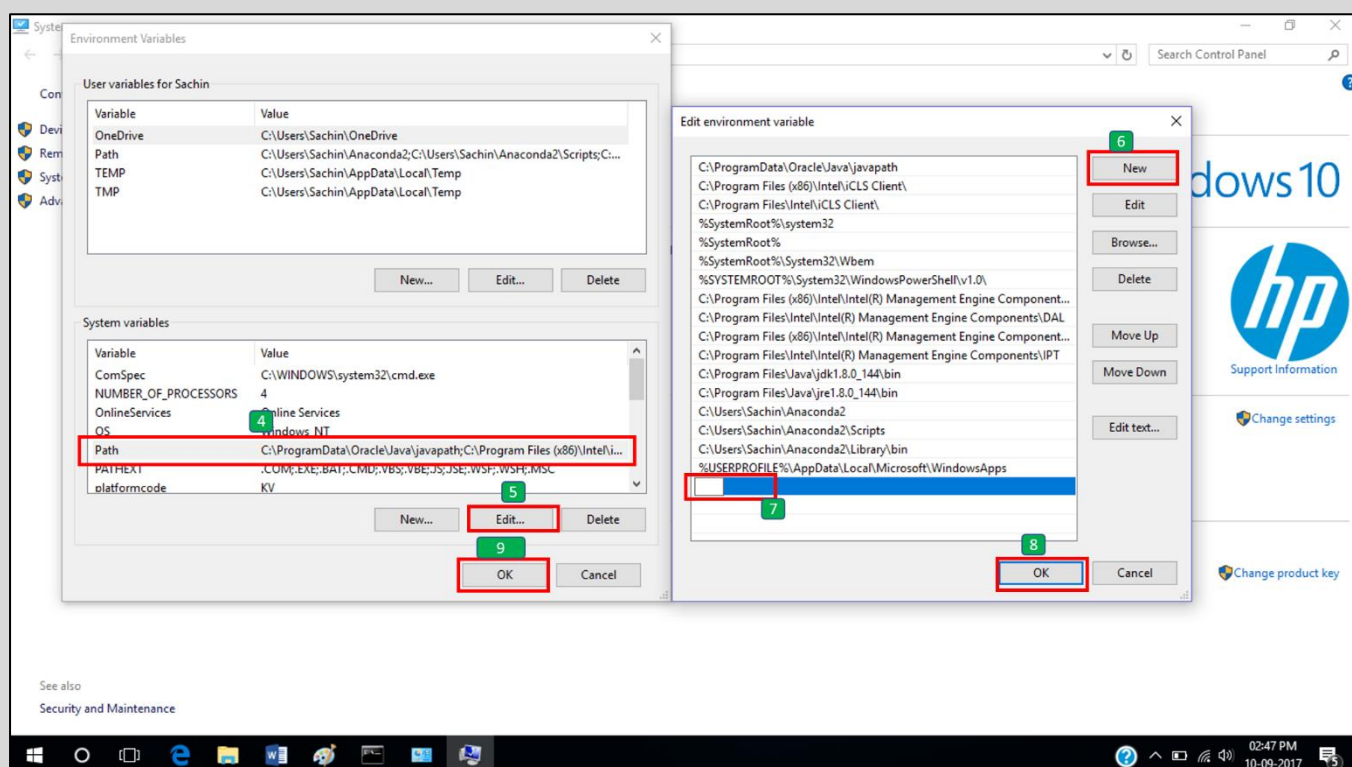


Figure 4

- Click on **New**
- Ctrl+V** to paste into step 7

Instruction - Paste your path without semicolon (;).

“C:\Program Files\Java\jdk1.8.0_144\bin”

Repeat same steps for jre click on **New** and **Ctrl+V** to paste the path.

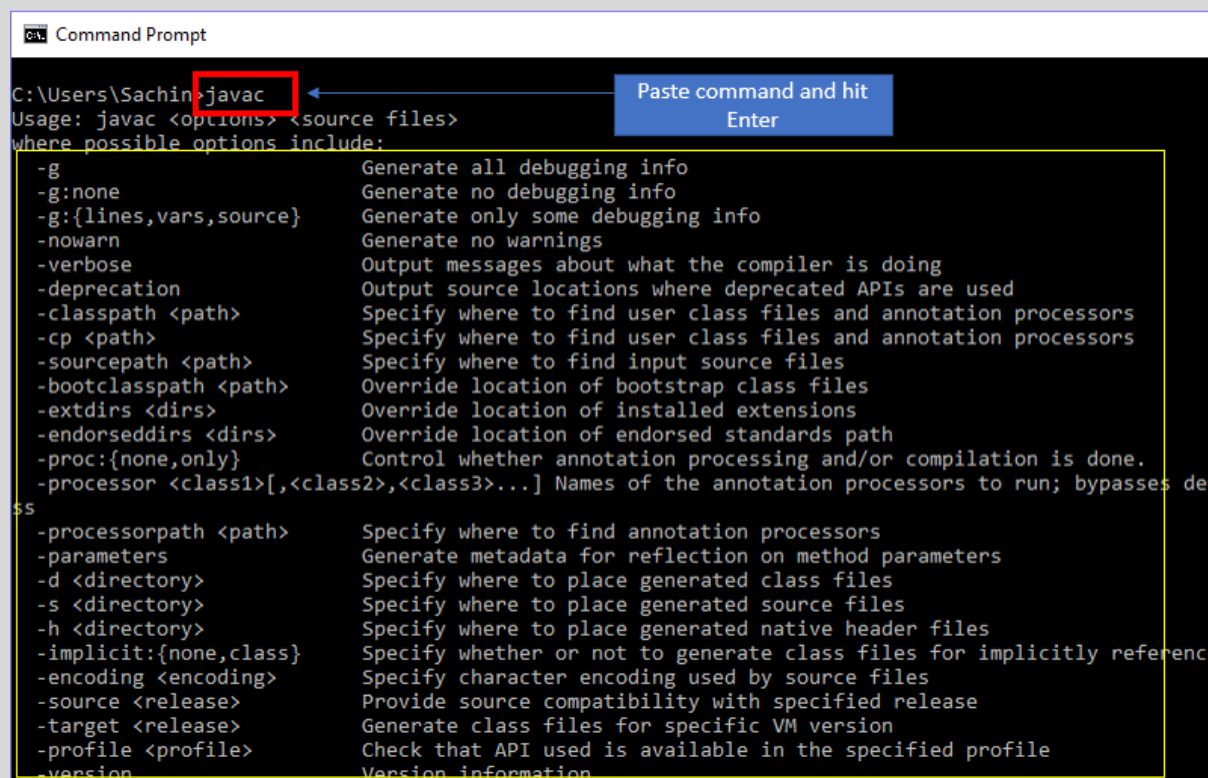
“C:\Program Files\Java\jre1.8.0_144\bin”

Then click **OK** and finish it.

8. Click **OK**

9. Click **OK**

To check java is successfully install or not type “**javac**” into command line if get a message as shown in Figure 5 then it is completed successfully. If not then check the path properly.



```
C:\Users\Sachin>javac
Usage: javac <options> <source files>
where possible options include:
  -g               Generate all debugging info
  -g:none          Generate no debugging info
  -g:{lines,vars,source} Generate only some debugging info
  -nowarn          Generate no warnings
  -verbose         Output messages about what the compiler is doing
  -deprecation     Output source locations where deprecated APIs are used
  -classpath <path> Specify where to find user class files and annotation processors
  -cp <path>       Specify where to find user class files and annotation processors
  -sourcepath <path> Specify where to find input source files
  -bootclasspath <path> Override location of bootstrap class files
  -extdirs <dirs>   Override location of installed extensions
  -endorseddirs <dirs> Override location of endorsed standards path
  -proc:{none,only} Control whether annotation processing and/or compilation is done.
  -processor <class1>[,<class2>,<class3>...] Names of the annotation processors to run; bypasses de
ss
  -processorpath <path> Specify where to find annotation processors
  -parameters       Generate metadata for reflection on method parameters
  -d <directory>    Specify where to place generated class files
  -s <directory>    Specify where to place generated source files
  -h <directory>    Specify where to place generated native header files
  -implicit:{none,class} Specify whether or not to generate class files for implicitly referenc
  -encoding <encoding> Specify character encoding used by source files
  -source <release> Provide source compatibility with specified release
  -target <release>  Generate class files for specific VM version
  -profile <profile> Check that API used is available in the specified profile
  -version          Version information
```

Figure 5

2.2 Anaconda

Follow this link <https://www.continuum.io/downloads> to download anaconda. Then install it on a computer by following installation step.

ChemSuite Tutorial

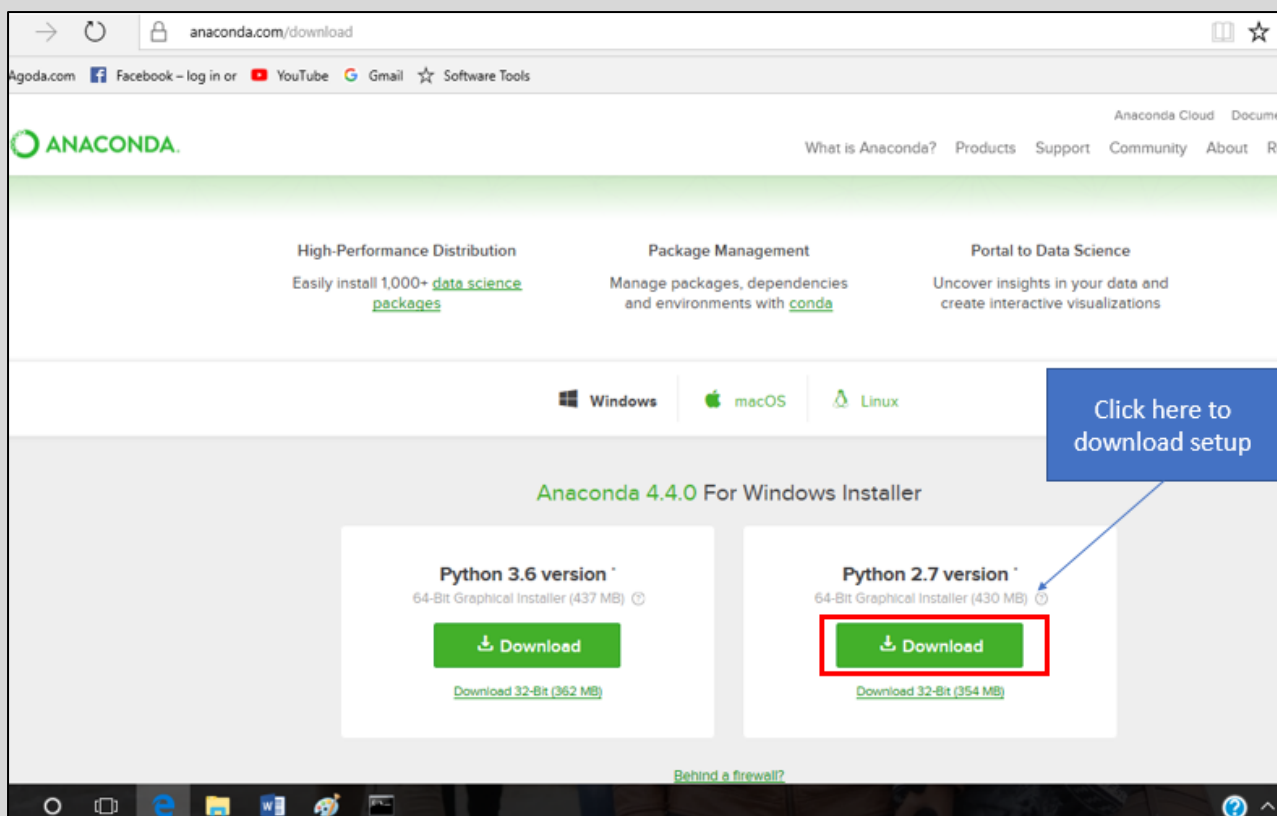


Figure 6

Type “**python**” into the command line, if it shows a message as shown in Figure 7 then follow steps given below to set path into the environment variable.

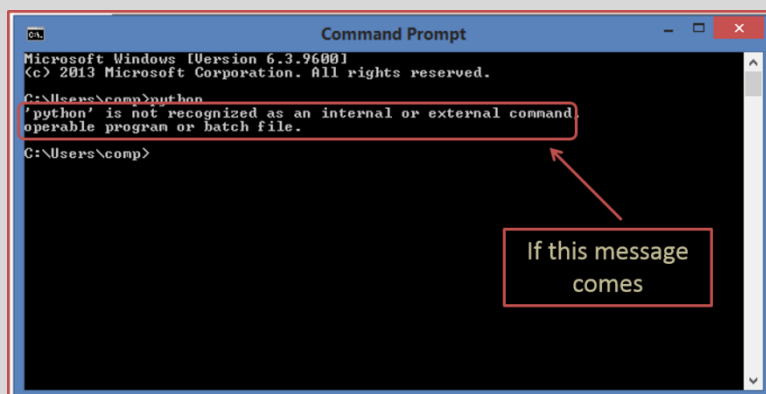


Figure 7

Steps to set path into environment variable -

1. Right click on my computer icon (**Properties**)
2. **Advanced system setting**
3. **Environment variable**
4. **Path**
5. **Edit**

Paste this path
“C:\Users\Sachin\Anaconda2;C:\Users\Sachin\Anaconda2\Scripts;C:\Users\Sachin\Anaconda2\Library\bin” replace **Sachin** with your computer name and click OK.

6. Click **OK**

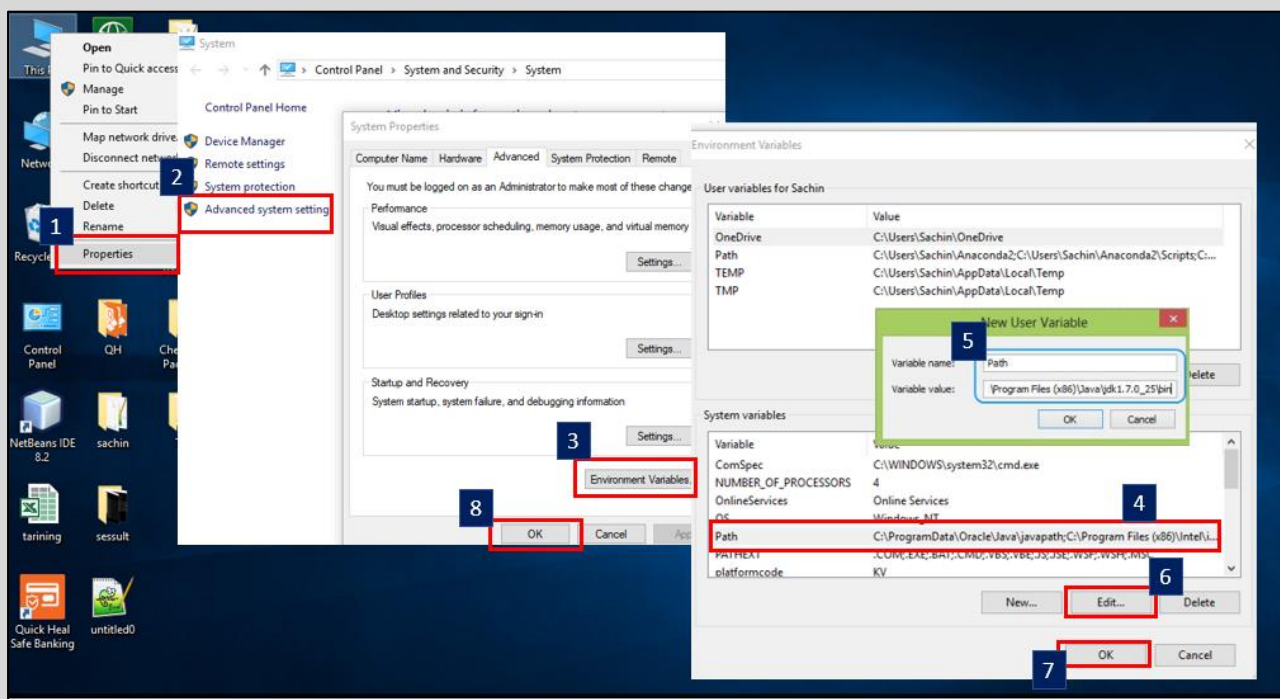


Figure 8

For windows 10 follow [Windows 10 path setting](#) given in the previous section and paste these paths one by one. Replace **Sachin** with your computer name.

C:\Users\Sachin\Anaconda2

C:\Users\Sachin\Anaconda2\Scripts

C:\Users\Sachin\Anaconda2\Library\bin

2.3 ChemSuite-Installer

ChemSuite-Installer is for a user who wants to install PyDPI, VC++, RDKit, Pytz, Scikit-learn and ChemSuite. If a user already has all pre-requisite package then the user needs to install ChemSuite only from ChemSuite-Installer. To download “ChemSuite-installer” go to <http://14.139.57.41/chemsuiteweb/download.php> or and download ChemSuite-Installer to install ChemSuite package. Extract the zip file and open it, then locate “**Installer**” as shown in Figure 9 below.

ChemSuite Tutorial

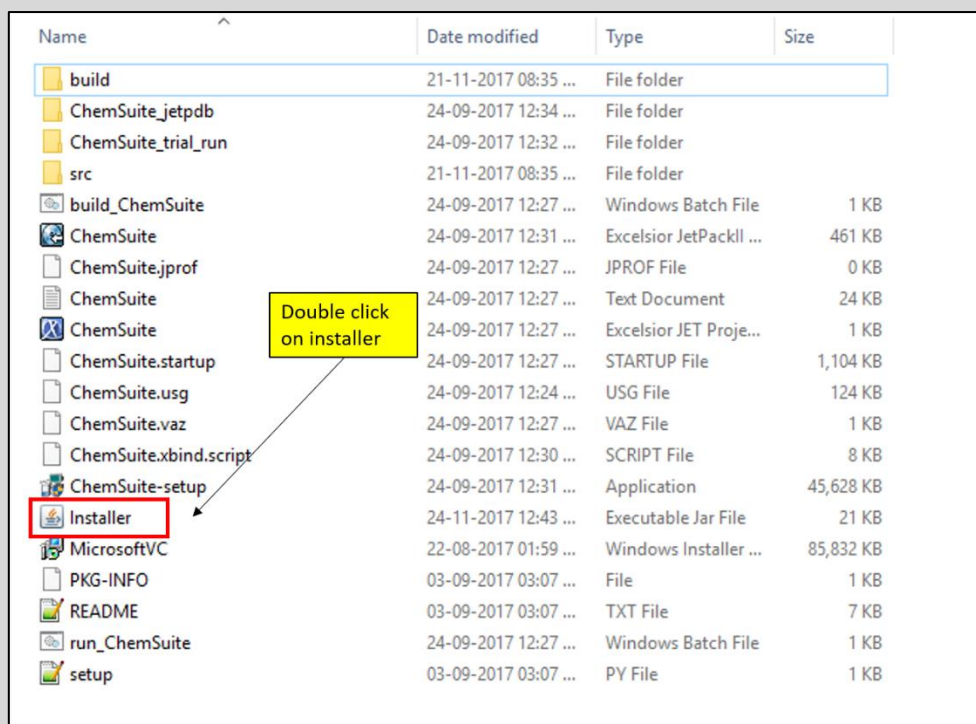


Figure 9

The Installer will open as shown in Figure 10 below just click on PyDPI, VC++, RDkit, Pytz, Scikit-learn and ChemSuite one by one to install all packages. If a user already has a package, so a user can skip that step and move to the next. All the package installations are independent of each other.

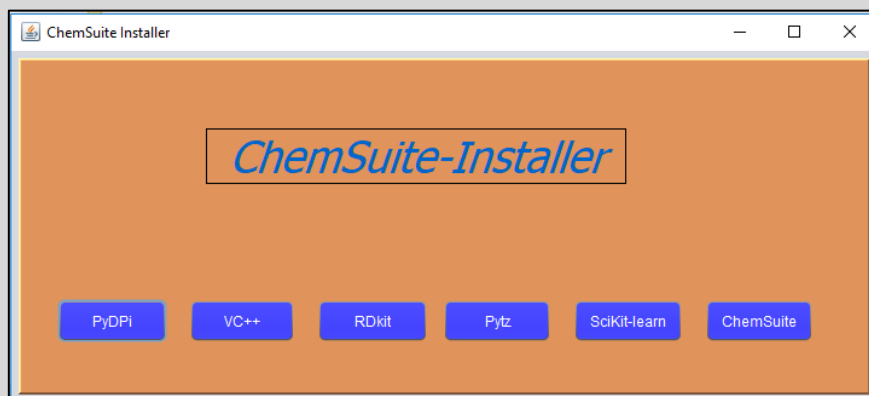
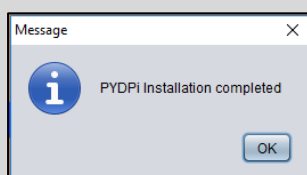


Figure 10

Follow steps to use ChemSuite Installer -

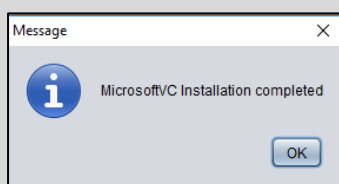
Step 1. Click on PyDPI button to start the installation.



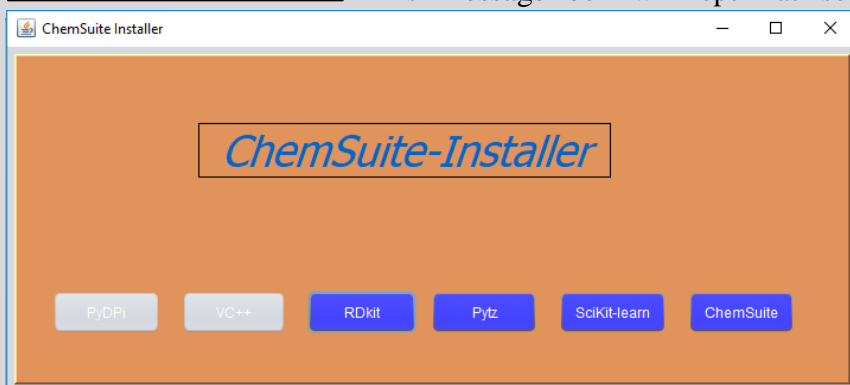
This message box will open as soon as installation complete.

Step 2. Click on MicrosoftVC button to start the installation.

ChemSuite Tutorial

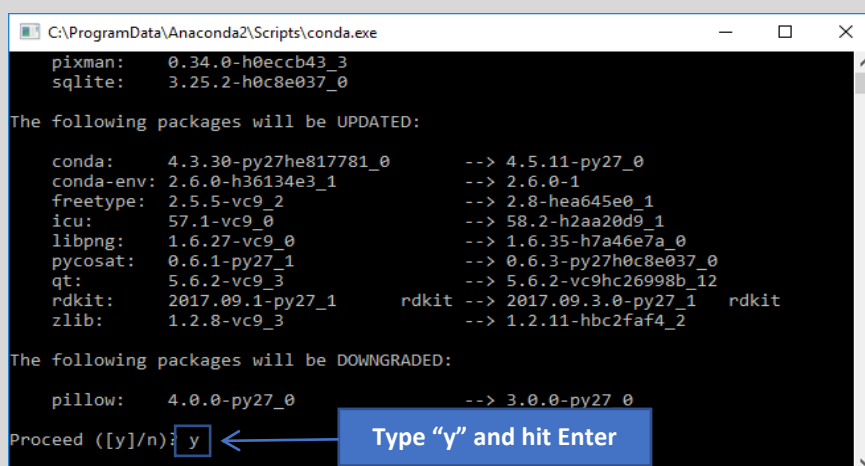


This message box will open as soon as installation complete.

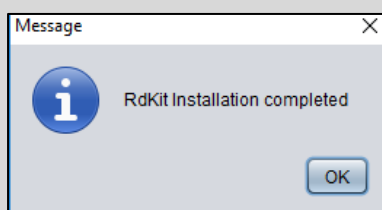


Buttons will get disabled after completion of the installation to avoid further overlaps of packages.

Step 3. Click on RDKit button to start the installation.



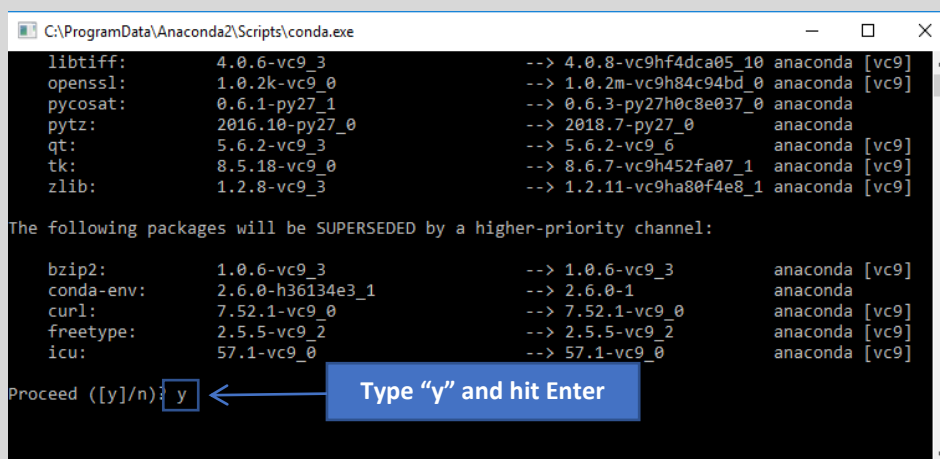
User needs to type “y” and hit enter to start installation process.



This message box will open as soon as installation complete.

Step 4. Click on Pytz button to start the installation.

ChemSuite Tutorial



```
C:\ProgramData\Anaconda2\Scripts\conda.exe

libtiff:      4.0.6-vc9_3      --> 4.0.8-vc9hf4dca05_10 anaconda [vc9]
openssl:     1.0.2k-vc9_0     --> 1.0.2m-vc9h84c94bd_0 anaconda [vc9]
pycosat:     0.6.1-py27_1     --> 0.6.3-py27h0c8e037_0 anaconda
pytz:        2016.10-py27_0   --> 2018.7-py27_0      anaconda
qt:          5.6.2-vc9_3      --> 5.6.2-vc9_6      anaconda [vc9]
tk:          8.5.18-vc9_0     --> 8.6.7-vc9h452fa07_1 anaconda [vc9]
zlib:        1.2.8-vc9_3      --> 1.2.11-vc9ha80f4e8_1 anaconda [vc9]

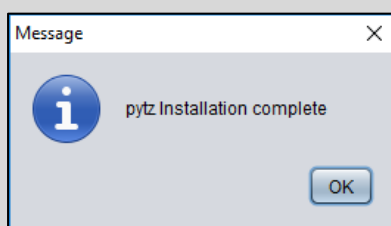
The following packages will be SUPERSEDED by a higher-priority channel:

bzip2:        1.0.6-vc9_3      --> 1.0.6-vc9_3      anaconda [vc9]
conda-env:    2.6.0-h36134e3_1 --> 2.6.0-1        anaconda
curl:         7.52.1-vc9_0     --> 7.52.1-vc9_0     anaconda [vc9]
freetype:     2.5.5-vc9_2     --> 2.5.5-vc9_2     anaconda [vc9]
icu:          57.1-vc9_0      --> 57.1-vc9_0      anaconda [vc9]

Proceed ([y]/n): y
```

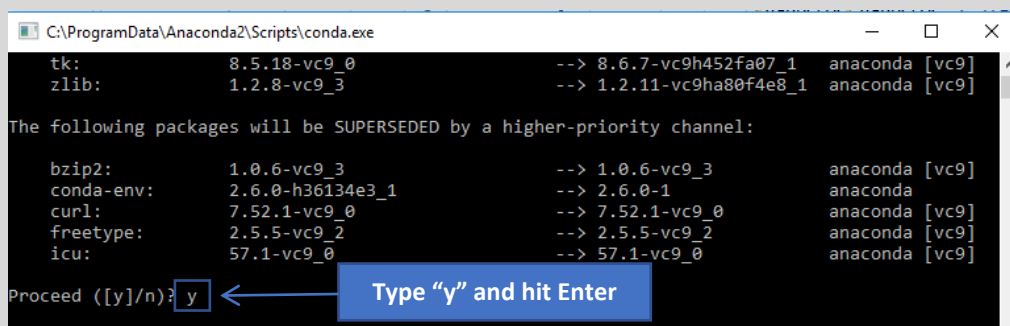
Type "y" and hit Enter

The user needs to type "y" and hit enter to start the installation process.



This message box will open as soon as installation complete.

Step 5. Click on SciKit-learn button to start the installation.



```
C:\ProgramData\Anaconda2\Scripts\conda.exe

tk:          8.5.18-vc9_0     --> 8.6.7-vc9h452fa07_1 anaconda [vc9]
zlib:        1.2.8-vc9_3      --> 1.2.11-vc9ha80f4e8_1 anaconda [vc9]

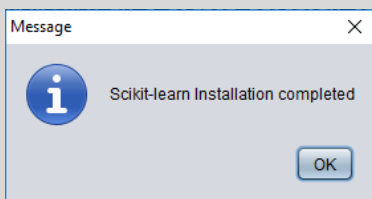
The following packages will be SUPERSEDED by a higher-priority channel:

bzip2:        1.0.6-vc9_3      --> 1.0.6-vc9_3      anaconda [vc9]
conda-env:    2.6.0-h36134e3_1 --> 2.6.0-1        anaconda
curl:         7.52.1-vc9_0     --> 7.52.1-vc9_0     anaconda [vc9]
freetype:     2.5.5-vc9_2     --> 2.5.5-vc9_2     anaconda [vc9]
icu:          57.1-vc9_0      --> 57.1-vc9_0      anaconda [vc9]

Proceed ([y]/n): y
```

Type "y" and hit Enter

The user needs to type "y" and hit enter to start the installation process.

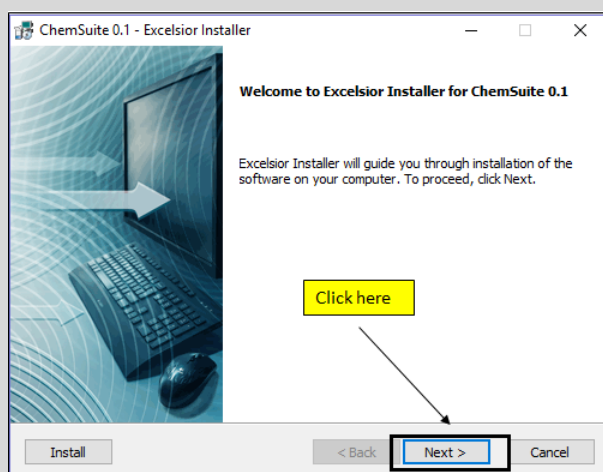


This message box will open as soon as installation complete.

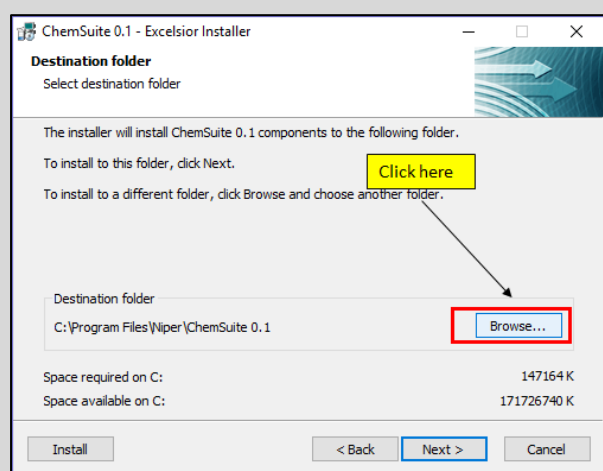
Step 6. Click on ChemSuite button to start the installation.

Click next to start

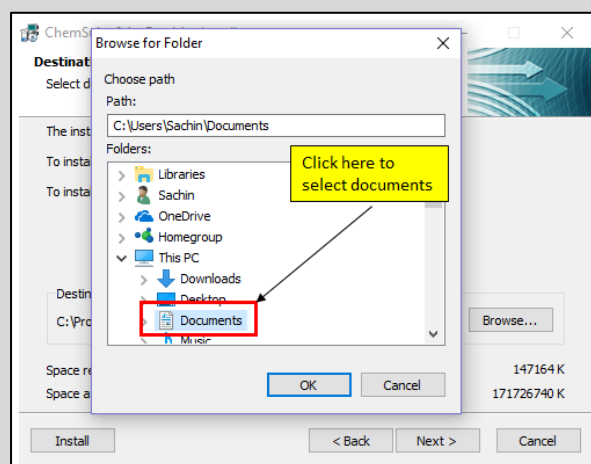
ChemSuite Tutorial



Click the browse button to select the directory

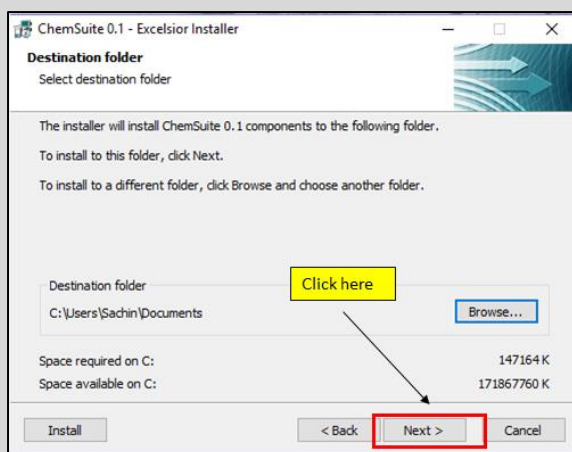


Select “**Document**” directory to install ChemSuite

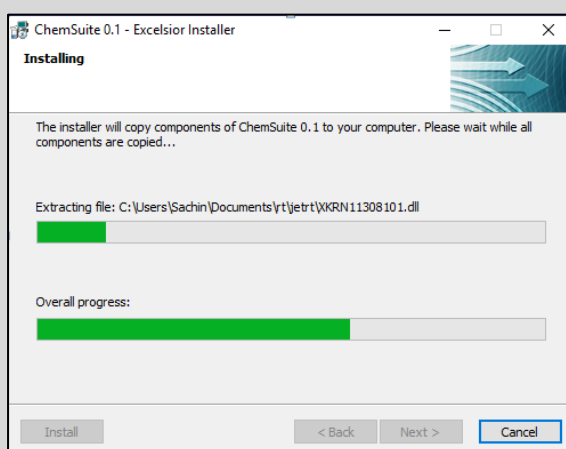


After selecting click the next button

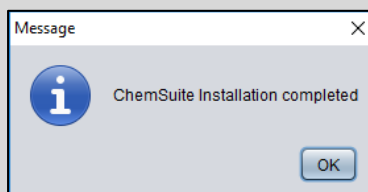
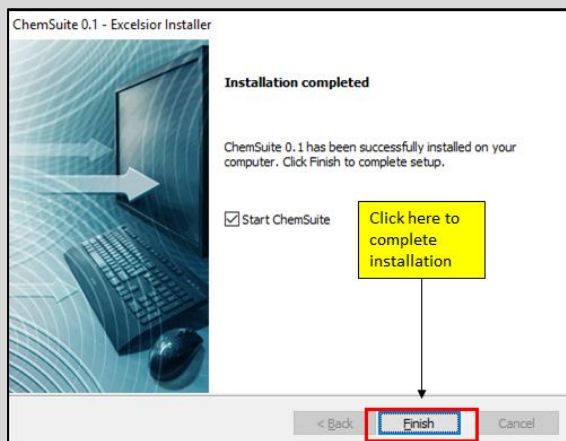
ChemSuite Tutorial



Wait for the completion of the process.



Click finish to complete the installation.



This message box will open as soon as installation complete.

3 Installation steps for Linux user

A user needs to install pre-requisite packages to use ChemSuite package in the Linux system.

Table 2 List of packages for ChemSuite

S/N	Package	Source
1.	Java (9.0.1)	http://www.oracle.com/technetwork/java/javase/downloads/index.html
2.	Anaconda (5.0.1)	https://www.anaconda.com/download/#linux
3.	PyDPI	https://pypi.python.org/pypi/conda/3.17.0
4.	RDkit	https://anaconda.org/rdkit/rdkit
5.	Pytz	https://anaconda.org/anaconda/pytz
6.	Sci-kit learn	https://anaconda.org/anaconda/scikit-learn
7.	ChemSuite (0.1)	http://14.139.57.41/chemsuiteweb/download.php

A user needs to install these packages listed in Table 2 to use ChemSuite package, All supported version and sources are given in Table 2. The user needs to download “**Linux-setup.zip**” file from given link and unzip the folder. The user needs to run “**ChemSuite.jar**” file by double click on it as shown in Figure 11.

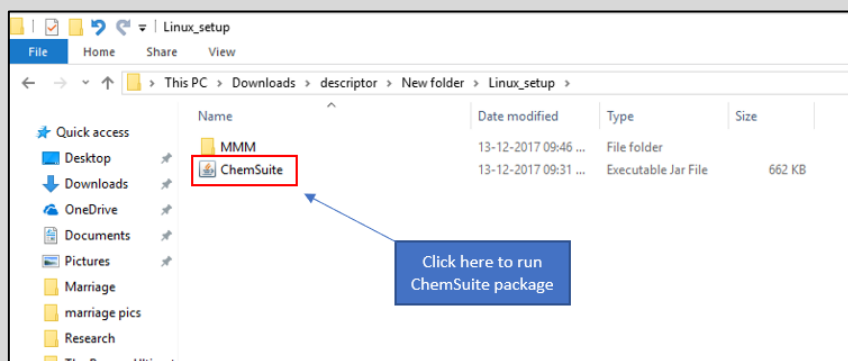


Figure 11

4 ChemSuite Tutorial

4.1 Home Page

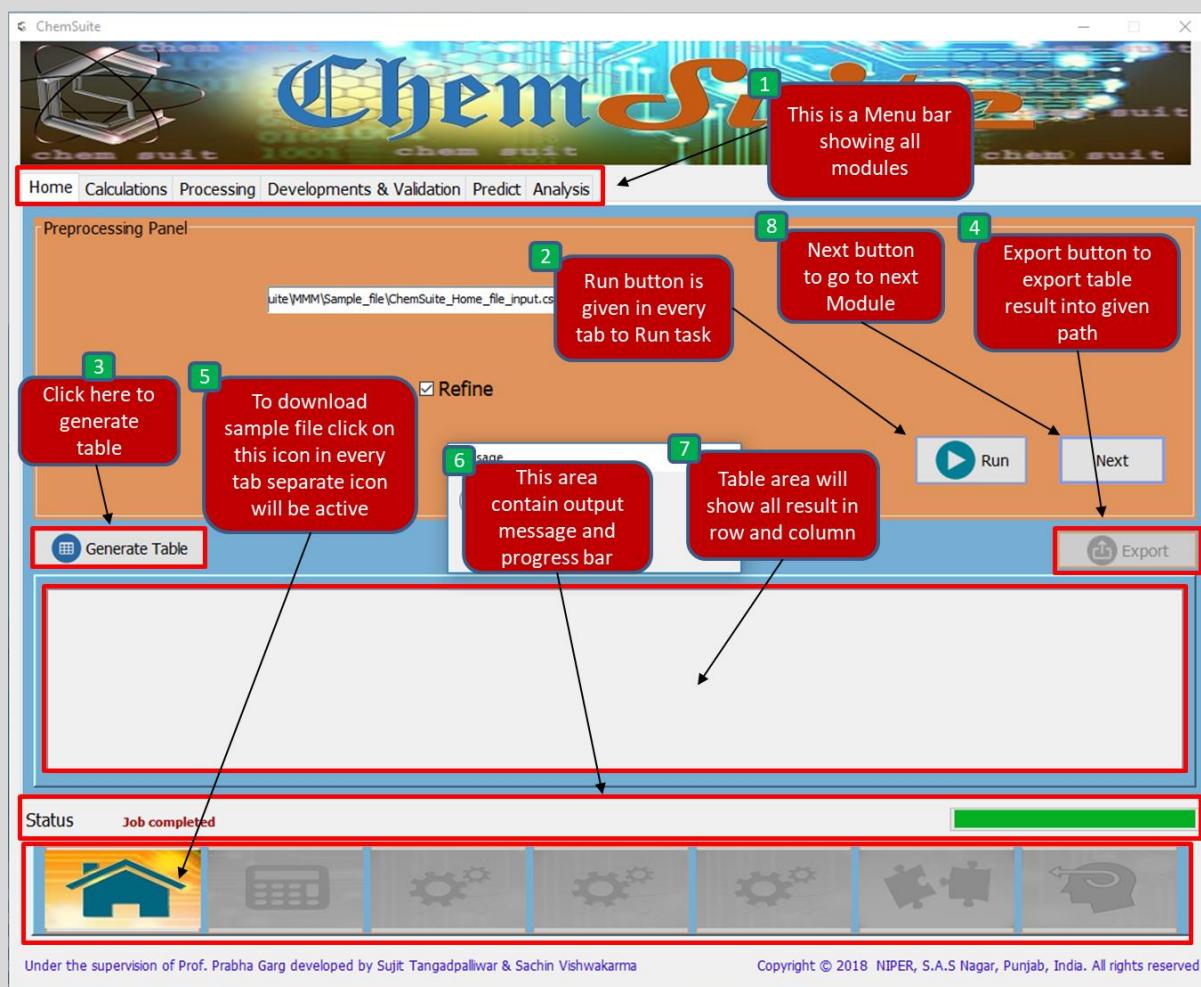
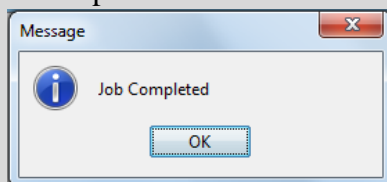


Figure 12

Home Image is showing common icon which will be same in all the tabs -

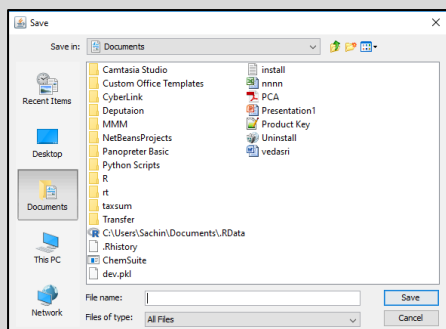
1. Step one contains all the modules name which user can choose to perform a particular task.
2. Step two showing the run button to perform a task.
3. Step three generate table button, On click table will be filled with data and ready to export. Generate table button only be active after run button



complete task this message will appear when job complete. Then generate table button will disable to avoid appending data into the same file.

4. Step four export button, On click user will see this file chooser panel

ChemSuite Tutorial



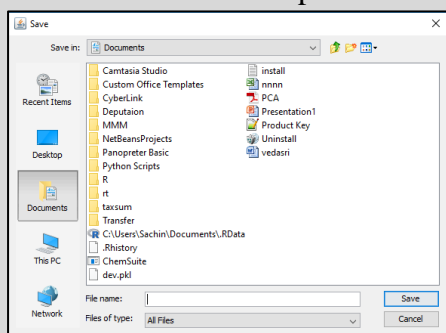
a user needs to choose a directory and write the name of a file then click save button to save file. This export button will be active after generate table button clicked.

5. Step five here indicating bottom panel like shown below at a time only one button



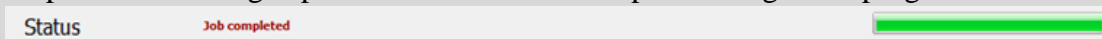
will be active for

example if a user is in predict **Predict** tab then only predict button will be active as shown in this example on click to this button file save panel will popup

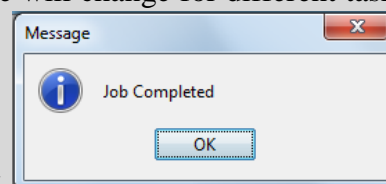


a user needs to choose a directory and write the name of a file then click save button to save samples file.

6. Step six is showing a panel which contains output message and progress bar where



a user can see when a process is completed. A message will change for different task



and task-bar will be filled 100% and one message box will appear on screen a user needs to click OK button.

7. Step seven is table area where all the result will be shown.




	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.01826824...	0.0	0.12126369...	0.17971236...	0.0	0.17971236...	1.0	0.12325575...	0.12326570...	0.12359550...	0.73994456...	0.13776729...	0.0
1	0.75991150...	0.0	0.52794075...	0.62933565...	0.86233978...	0.62933565...	0.13766021...	0.51318166...	0.51305915...	0.51685393...	0.55127096...	0.36685554...	0.7
2	0.0	0.0	0.0	0.37078960...	0.0	0.629294939...	0.0	0.0	0.0	0.0	1.0	0.82435307...	0.0
3	1.0	0.0	0.77132025...	0.76777543...	0.75687930...	0.76777543...	0.24312069...	0.74309071...	0.74386009...	0.69662921...	0.18088777...	0.20302939...	1.0
4	0.70790986...	0.0	0.31178509...	0.21914151...	0.26571857...	0.21914151...	0.73428142...	0.30799279...	0.30807503...	0.28098887...	0.70547188...	0.0	0.3
5	0.67876415...	0.0	0.63194626...	0.85581911...	0.87152815...	0.85581911...	0.12847184...	0.65725068...	0.65717028...	0.71910112...	0.03013559...	0.97344007...	0.6
6	0.26530561...	0.0	0.71862652...	0.24225959...	0.93253629...	0.24225959...	0.06746370...	0.71875911...	0.71869750...	0.73033707...	0.62114828...	0.08516242...	0.2


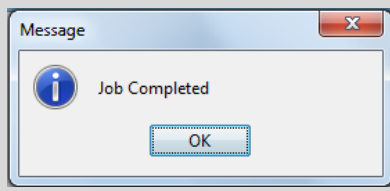
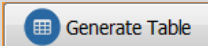
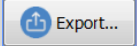
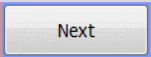

8. Step eight **Next** next button, On click a user will move to the next tab.

These icons will be the same in all module (Tab) once explained here.

4.2 Preprocessing Panel

The first click on Home Tab pre-processing panel will appear. In this panel just follow the steps as shown in Figure 13.

1. Click on  Tab.
2. Click on  button to browse .csv file which a user wants to refine (file format and column names should be same as given in sample file).
3. Select an option  Optimize or Refine or both of them as per user requirement.

4. Click  button  this message will come once job completed.
5. Click on  button to generate result.
6. Click on  button to export data into folder.
7. Click  button to go to next tab.
8. Click on  to save supported sample file format for refinement and optimization file into a user-defined directory.

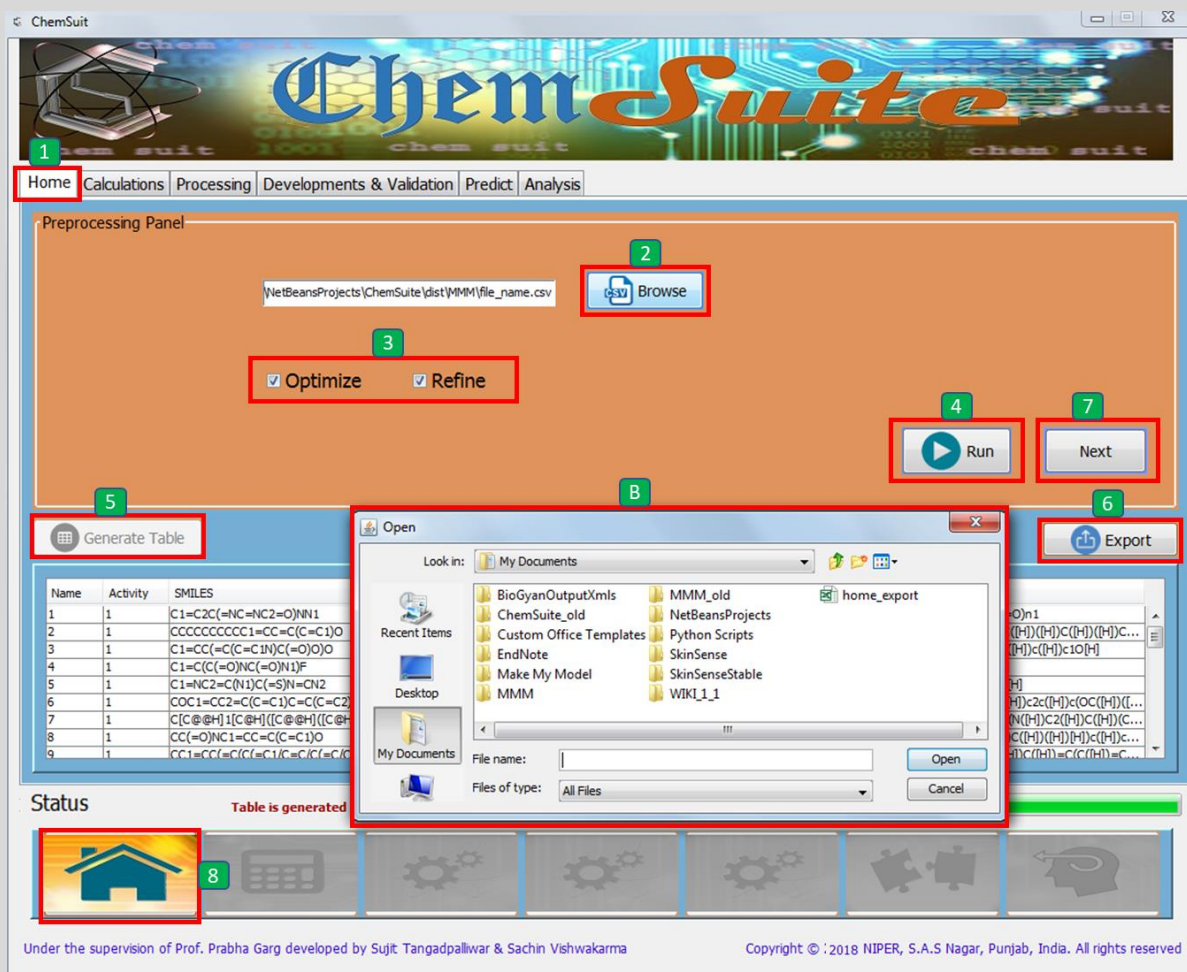


Figure 13

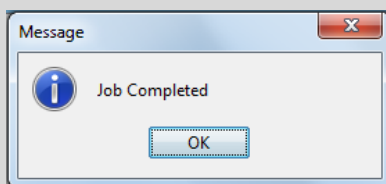
4.3 Calculation Tab

Follow the the steps for Calculation as shown in Figure 14.





1. Click on **Calculations** Tab.
2. Two options are shown here **Contineous with Previous** and **Import New File** to calculate descriptor. A user can either continue with the previous step which will take ChemSuite generated refine/optimized file for descriptor calculation or can **Browse** new refine/optimized file as per need (file format and column names should be same as given in sample file).
3. Select the type of descriptor user wants to calculate by just selecting **RDkit** **PyDPI** **PaDEL** RDkit, PyDPI or PaDEL one at a time. The user can also choose fingerprint or descriptor or both. On the basis of a user selection one single file will be generated containing all descriptors.
4. After selection click **Run** button to start calculating descriptor

```
C:\Program Files (x86)\Common Files\Oracle\Java\javapath\java.exe
Processing AUTOGEN_output_167 in output.smi (167/200). Average speed: 3.20 s/mol.
Processing AUTOGEN_output_168 in output.smi (168/200). Average speed: 3.19 s/mol.
Processing AUTOGEN_output_169 in output.smi (169/200). Average speed: 3.38 s/mol.
Processing AUTOGEN_output_170 in output.smi (170/200). Average speed: 3.37 s/mol.
Processing AUTOGEN_output_171 in output.smi (171/200). Average speed: 3.36 s/mol.
Processing AUTOGEN_output_172 in output.smi (172/200). Average speed: 3.38 s/mol.
Processing AUTOGEN_output_173 in output.smi (173/200). Average speed: 3.36 s/mol.
Processing AUTOGEN_output_174 in output.smi (174/200). Average speed: 3.34 s/mol.
Processing AUTOGEN_output_175 in output.smi (175/200). Average speed: 3.33 s/mol.
Processing AUTOGEN_output_176 in output.smi (176/200). Average speed: 3.31 s/mol.
Processing AUTOGEN_output_177 in output.smi (177/200). Average speed: 3.30 s/mol.
Processing AUTOGEN_output_178 in output.smi (178/200). Average speed: 3.29 s/mol.
Processing AUTOGEN_output_179 in output.smi (179/200). Average speed: 3.33 s/mol.
Processing AUTOGEN_output_180 in output.smi (180/200). Average speed: 3.31 s/mol.
KEYError: Unknown distance key in pSet: O=Csp2 take default bond length: 1.5
Processing AUTOGEN_output_181 in output.smi (181/200). Average speed: 3.30 s/mol.
Processing AUTOGEN_output_182 in output.smi (182/200). Average speed: 3.29 s/mol.
Processing AUTOGEN_output_183 in output.smi (183/200). Average speed: 3.29 s/mol.
Processing AUTOGEN_output_184 in output.smi (184/200). Average speed: 3.32 s/mol.
Processing AUTOGEN_output_185 in output.smi (185/200). Average speed: 3.32 s/mol.
Processing AUTOGEN_output_186 in output.smi (186/200). Average speed: 3.30 s/mol.
Processing AUTOGEN_output_187 in output.smi (187/200). Average speed: 3.28 s/mol.
Processing AUTOGEN_output_188 in output.smi (188/200). Average speed: 3.29 s/mol.
Processing AUTOGEN_output_189 in output.smi (189/200). Average speed: 3.30 s/mol.
Processing AUTOGEN_output_190 in output.smi (190/200). Average speed: 3.33 s/mol.
Processing AUTOGEN_output_191 in output.smi (191/200). Average speed: 3.30 s/mol.
Processing AUTOGEN_output_192 in output.smi (192/200). Average speed: 3.28 s/mol.
Processing AUTOGEN_output_193 in output.smi (193/200). Average speed: 3.27 s/mol.
```

The command line will popup where all the process will be done after that “Job completed”



a message will come once job completed.

5. Click  **Generate Table** button to generate a table where a user can see data.
6. Click  **Export...** button to export data that contain all descriptor into a given path.
7. Click  **Next** button to go to the next tab.
8. Click on  to save supported sample file format for descriptor calculation file into the user-defined directory.

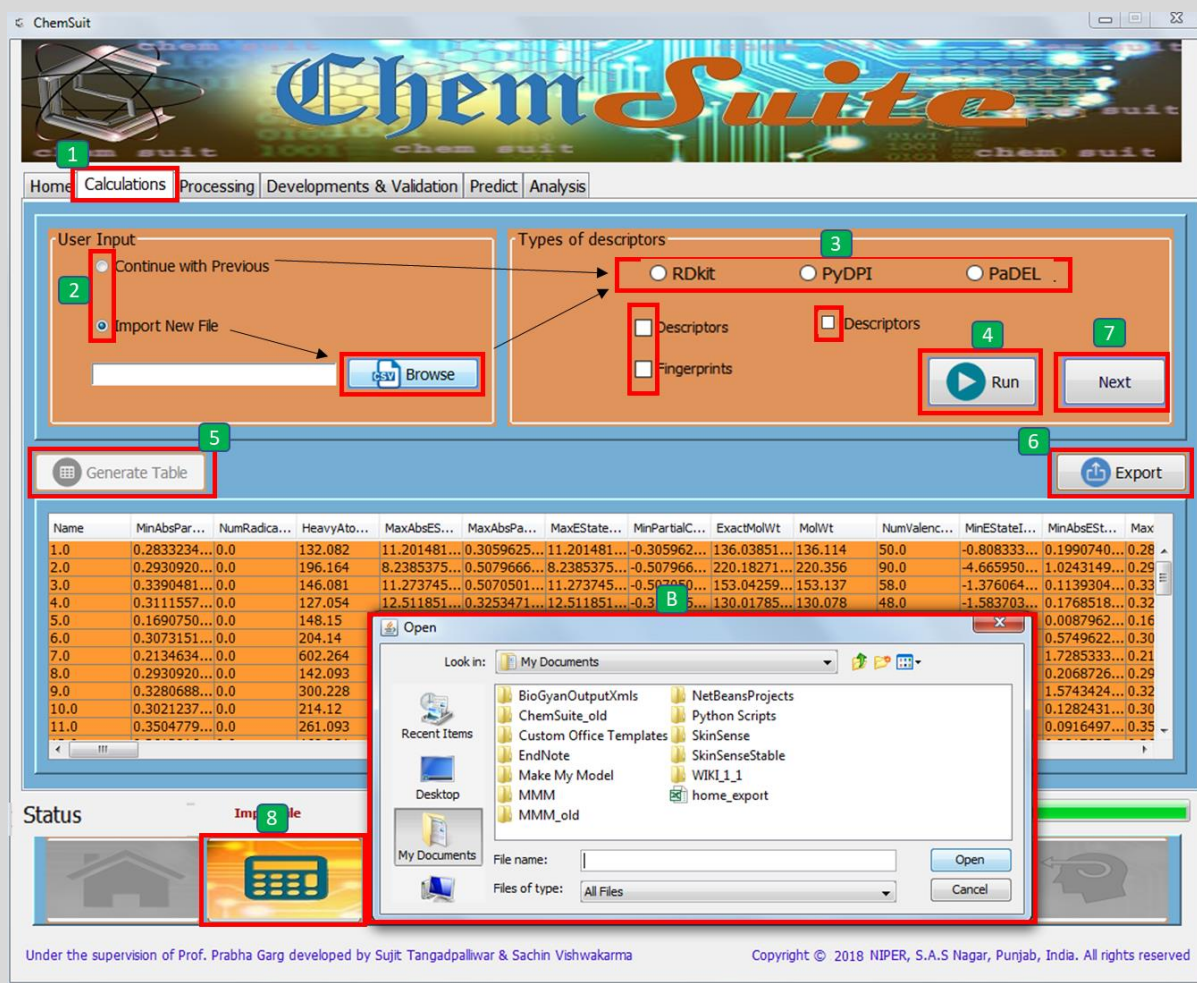


Figure 14

4.4 Processing Tab


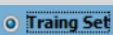
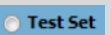
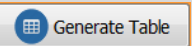
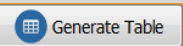

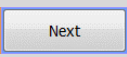

Follow the steps for splitting the dataset into training and test set as shown in Figure 15.

1. Click **Processing** Tab.
2. Next click **Splitting** Tab.
3. Two options are given **Continue with Previous** and **Import New File**. A user needs to select by click on the radio button. If a user select continue with previous then it will take previously generated descriptor file and will continue splitting. If a user select import new file then a user will have to browse **Browse** descriptor file here to continue (file format and column names should be same as given in sample file).
4. On the basis of previous selection next step is to enter splitting ratio for training and test set

Training	Test
80	20

 20:80 or 30:70.

ChemSuite Tutorial

- Click  button to start splitting data set this message will come once job completed.
- Step six   only one at a time can be selected and then click on  only after selecting the train set or test set.
-  button will be active click on it to generate table.
- Click on  button to export result into the given path.
- Click  button to go to the next tab.
- Click on  to save supported sample file format for splitting data file into the user-defined directory.

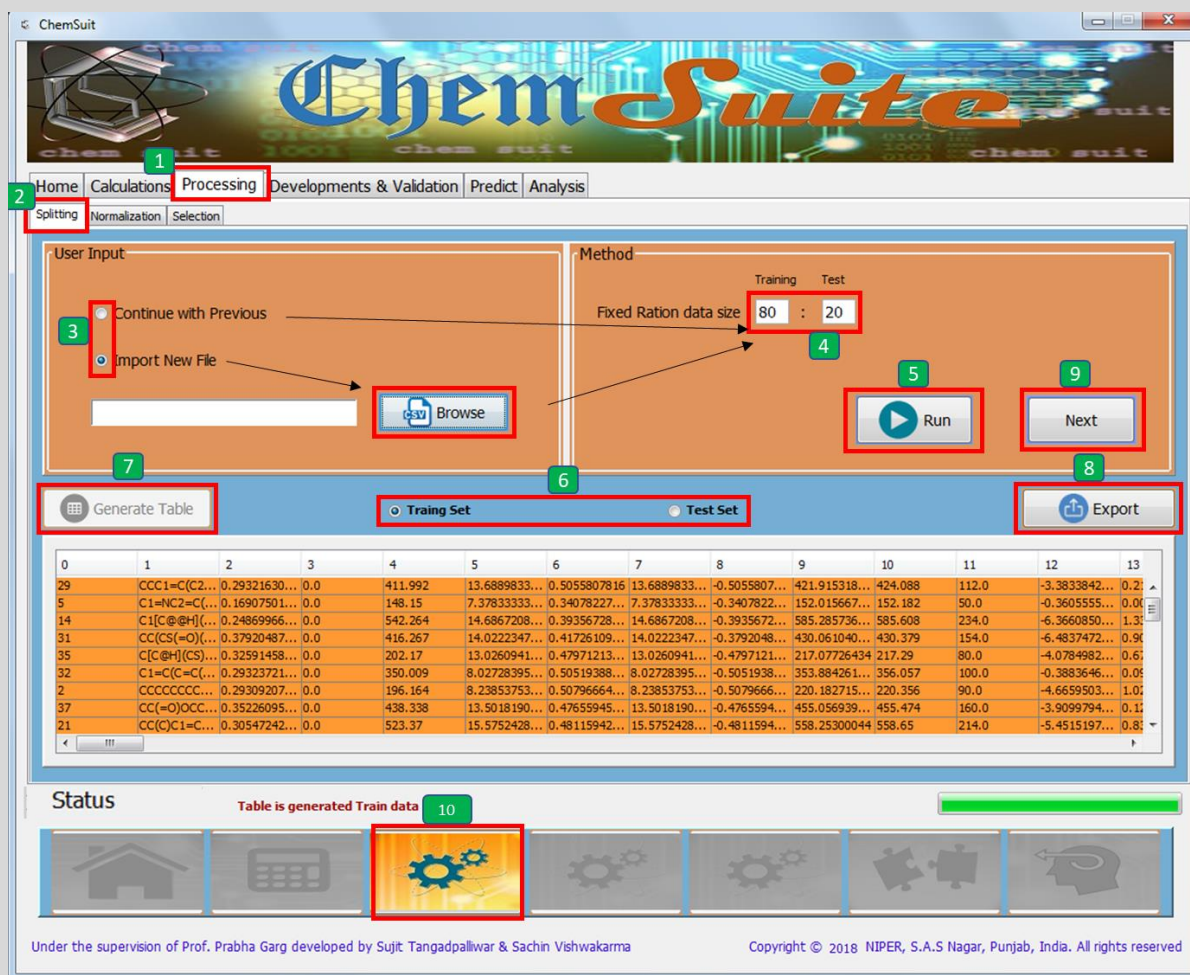
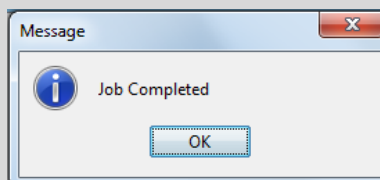
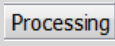
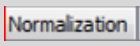
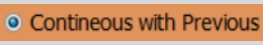
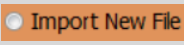

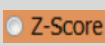
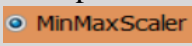


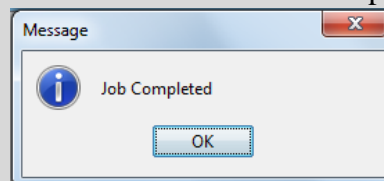
Figure 15


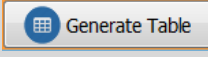
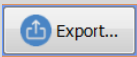
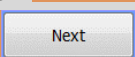

4.5 Normalization

Follow these steps for the Normalization as shown in Figure 16.

ChemSuite Tutorial

1. Click  Tab.
2. Next click  Tab.
3. Two options are given  and . A user needs to select by click on the radio button. If a user select continue with previous then it will take the previous generated splitted file and will continue normalization. If user select import new file then the user will have to browse  splitted file here to continue (file format and column names should be same as given in sample file).
4. On the basis of previous selection next step is to select the normalization method  or . On the basis of the selection action will be performed.



5. Click  button to normalize data. This message will come once job completed.
6. Click on  button to show data into a table format.
7. Click on  button to export result into the given path.
8. Click  button to go to the next tab.
9. Click on  to save supported sample file format for normalization file into the user-defined directory.

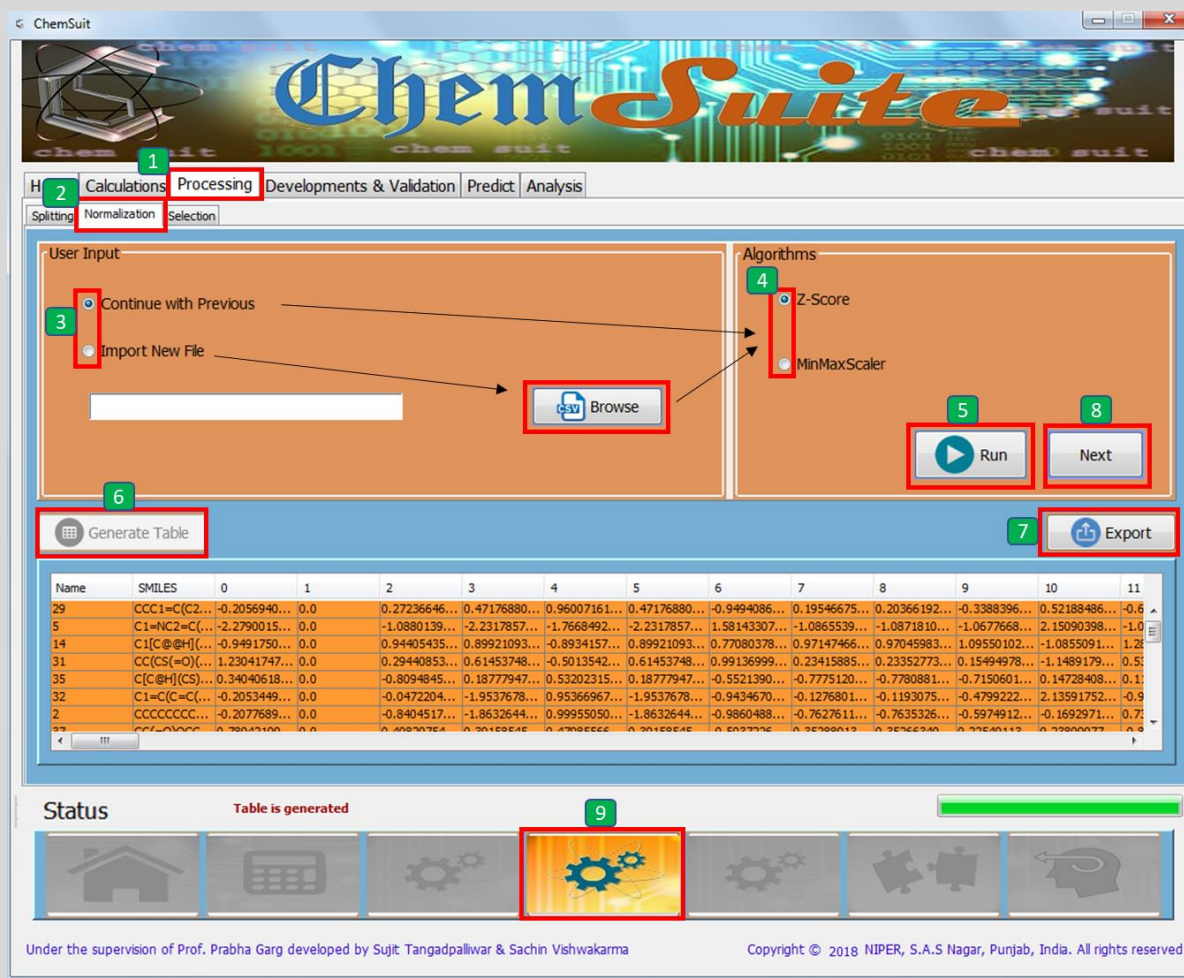



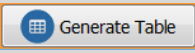
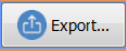

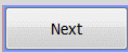
Figure 16

4.6 Feature selection

Follow these steps for Feature selection as shown in Figure 17.

1. Click **Processing** Tab.
2. Next click **Selection** Tab.
3. Two options are given **Continue with Previous** and **Import New File**. A user needs to select by click on the radio button. If a user select continue with previous then it will take the previous generated normalized file and will continue feature selection. If a user select import new file then user will have to browse **Browse** normalized file here to continue (file format and column names should be same as given in sample file).
4. On the basis of previous selection next step is to select an algorithm from given option Chi2, RF, REF, VR. On the basis of your selection parameter will be shown the user can change according to their need. Example for the Chi2 user can define K value and all other parameters.
5. Step five to select the parameter for the different method.

ChemSuite Tutorial

6. Click  button to normalize data. This message will come once job completed.
7. Click on  button to show data into a table format.
8. Click on  button to export result into given path.
9. Click on  to save supported sample file format for feature selection file into the user-defined directory.
10. Click  button to go to the next tab.

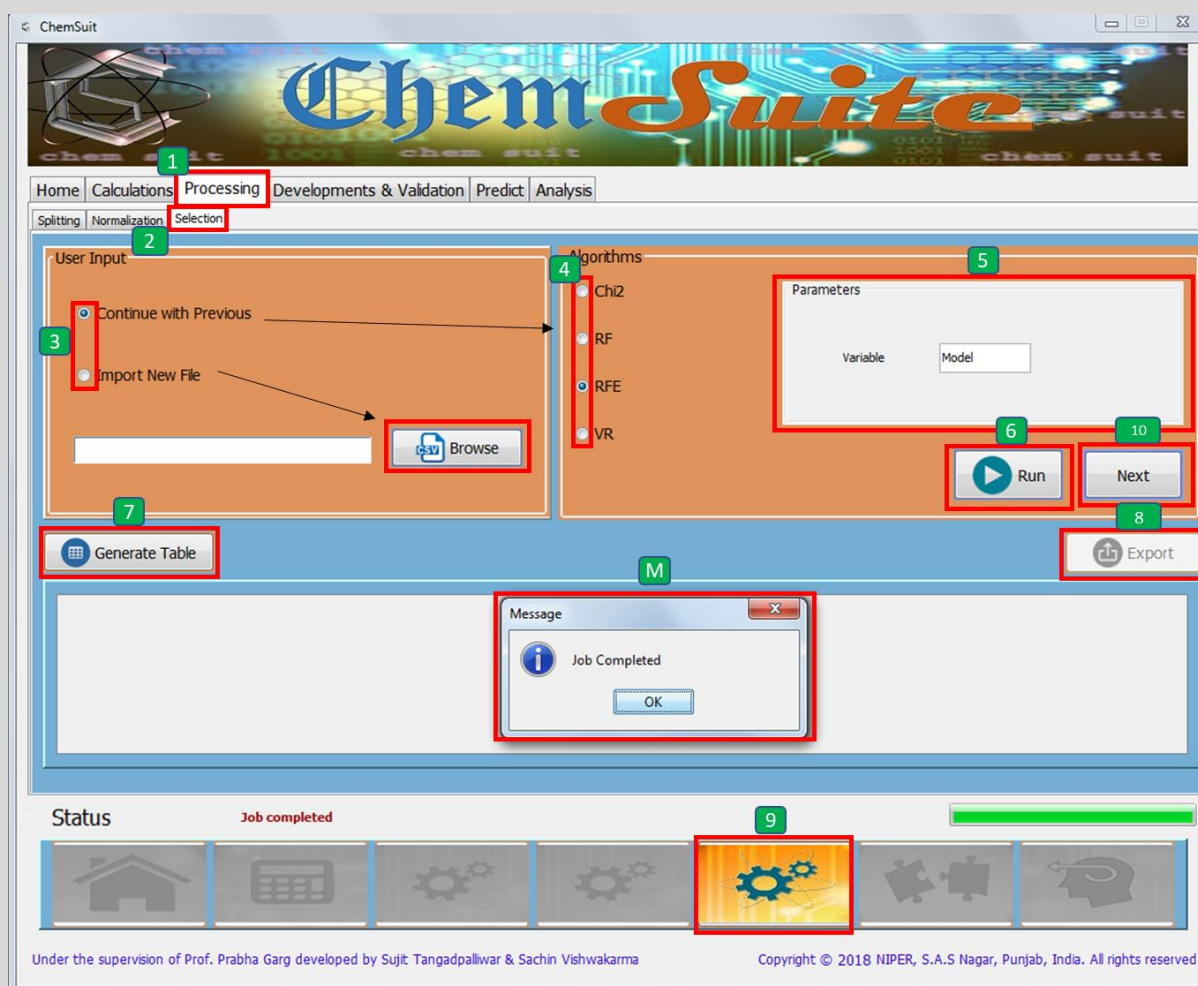
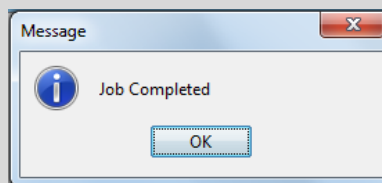



Figure 17


4.7 Development & Validation

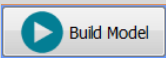
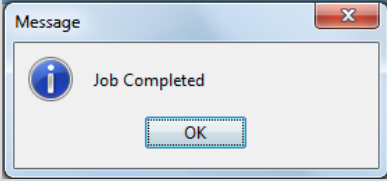
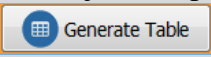
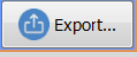
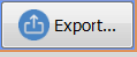
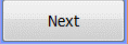

Follow these the steps for development and validation as shown in Figure 18.

1. Click **Developments & Validation** Tab.
2. Two options are given **Continuous with Previous** and **Import New File**. A user needs to select by click on the radio button. If a user select continue with previous then it will take

ChemSuite Tutorial

the previous generated feature selected file and will continue model development. If a user select import new file then the user will have to browse  feature selected file here to continue (file format and column names should be same as given in sample file).

3. On the basis of previous selection next step is to select the model development method  like Random forest, Support vector machine etc.
4. Select the parameter for example for Random forest set parameter like random state, number of the estimator, number of fold etc as per requirement or can leave default.
5. Select ☒ Training whole dataset or ☐ Cross Validation

6. Click  button to build model  this message will come once job completed.
7. Click on  button to show data into a table format.
8. Click on  button to export result into the given path.
9. Click on  button to export .pkl file into the given path.
10. Click  button to go to the next tab.
11. Click on this  to save supported sample file format for model building file into the user-defined directory.

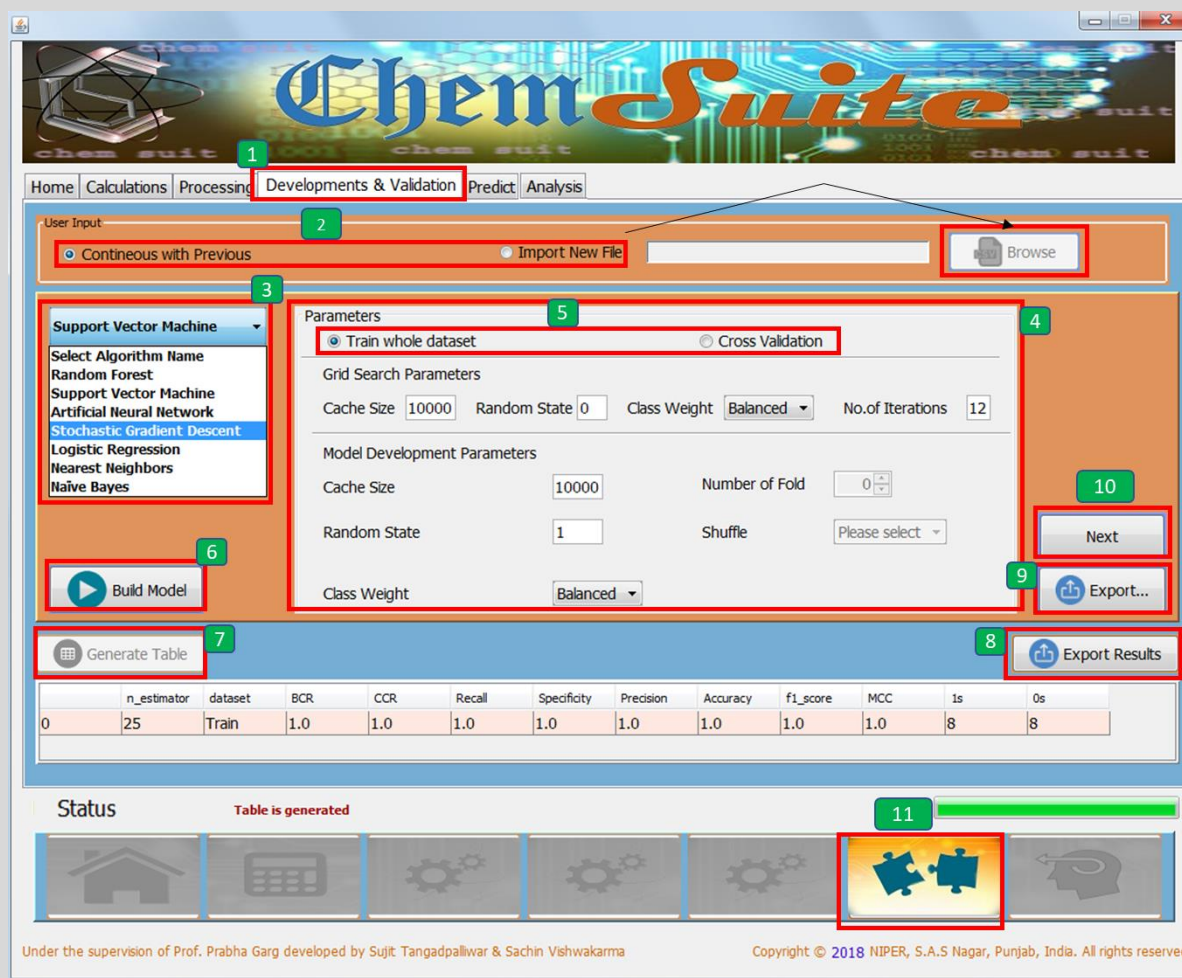
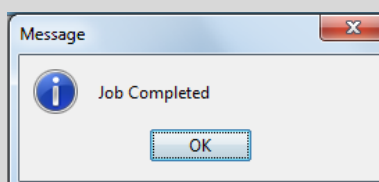


Figure 18

4.8 Predict

Follow these the steps for the Prediction as shown in Figure 19.

1. Click **Predict** Tab.
2. Two options are given **Continuous with Previous** and **Import New File**. A user needs to select by click on the radio button. If a user select continue with previous then it will take previously generated test file and will predict. If a user select import new file then the user will have to browse **Browse** already built model (.pkl file) and a test file (.csv file).
3. Click on **Predict** button to predict this message will come once job completed.
4. Click on **Generate Table** button to show data into a table format.
5. Click on **Export...** button to export result into the given path.



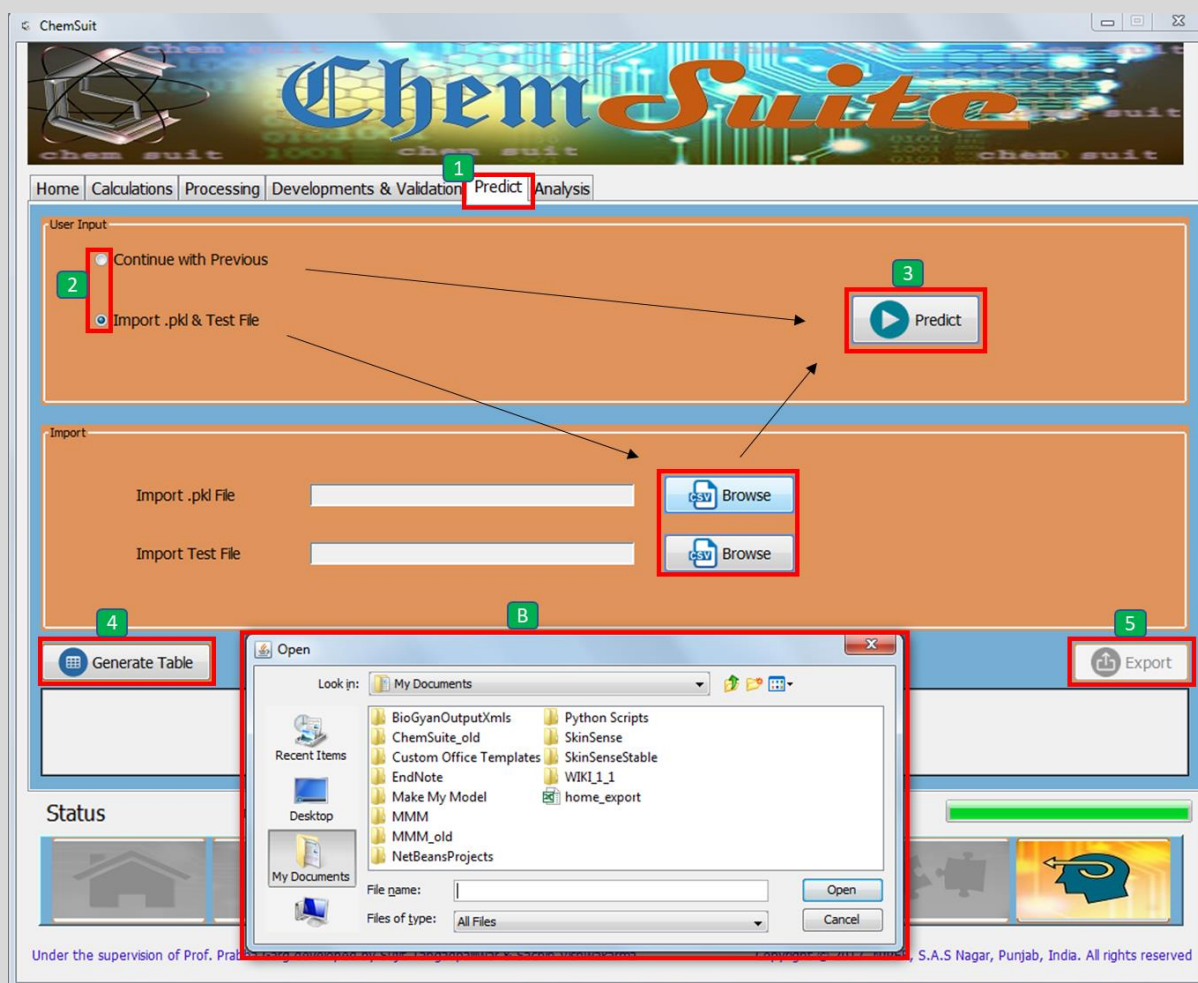


Figure 19

4.9 Analysis

Follow these steps for the Analysis as shown in Figure 20.

1. Click on **Analysis** tab.
2. Two option are given **Training Set** **Test Set** select one option.
3. Click on **Generate table** to view data into a table.
4. Click on **Export...** button to export result into the given path.

The screenshot shows the ChemSuite application window. The 'Analysis' tab is selected, indicated by a green box labeled '1'. The main area displays a table titled 'Actual value and Predicted value table'. To the right of the table are two radio buttons for 'Training set' (selected) and 'Test set', with a green box labeled '2' around them. Below these are two buttons: 'Generate table' (labeled with a green box '3') and 'Export' (labeled with a green box '4').

Name	SMILES	Actual_activity	Predicted_activity
Nicotine	<chem>CN1CCC[C@H]1C1=CC(=O)OC2=CC=CC=C2</chem>	0	0
4-methylumbellifero...	<chem>CC1=CC(=O)OC2=CC=CC=C2</chem>	0	0
Pluronic L61	<chem>C1CO1.CC1CO1</chem>	1	1
sorafenib	<chem>C1=CC(=O)C1=NC=CC=C1</chem>	1	1
Cimetidine	<chem>CN=C(NCSCC1=C(C)CC1)</chem>	0	0
Lovastatin acid	<chem>CC[C@H](C)C(=O)O</chem>	0	0
Hoechst 33342	<chem>CCOC1=CC=C(C=C1)C</chem>	1	1
Zidovudine	<chem>CC1=CN([C@H]2C[C@@H](C2)N)C1</chem>	0	0
Atorvastatin lactone	<chem>CC(C)C1=C(C(=O)O)CC1</chem>	1	1
Imatinib mesylate	<chem>CS(O)(=O)=O.CN1=CC=C(C=C1)C</chem>	1	1
tramadol	<chem>COC1=CC=CC(=C1)C</chem>	0	0
Quinine	<chem>COC1=CC2=C(C=C1)C</chem>	1	1
Diethylstilbestrol	<chem>CCC(=C(C)C)C1=CC=CC=C1</chem>	1	1
omeprazole	<chem>COC1=CC=C2N=C(C=C2)C1</chem>	0	0
Ochratoxin A	<chem>C[C@H]1CC2=C(C=C1)C</chem>	1	1
Simvastatin acid	<chem>CCC(C)C(C(=O)O)C</chem>	0	0

Figure 20

Thank you for using ChemSuite