To install the package from a local machine use the code below. Note that the packages lme4 and stringr are also necessary and need to be installed.

```r
#install.packages("modSel_0.1.0.tar.gz", repos = NULL, type = "source")
library(modSel)
```

**Example 1**

We will illustrate the use of the model selection algorithm with *Boston* data set that can be found in the MASS library.

```r
library(MASS)
head(Boston)
```

```
## crim  zn  indus chas  nox  rm age  dis  rad  tax  pptratio  black
## 1 0.00632 18 2.31 0 0.538 6.575 65.2 4.0900 1 296 15.3 396.90
## 2 0.02731 0 7.07 0 0.469 6.421 78.9 4.9671 2 242 17.8 396.90
## 3 0.02729 0 7.07 0 0.469 7.185 61.1 4.9671 2 242 17.8 392.83
## 4 0.03237 0 2.18 0 0.458 6.998 45.8 6.0622 3 222 18.7 394.63
## 5 0.06905 0 2.18 0 0.458 7.147 54.2 6.0622 3 222 18.7 396.90
## 6 0.02985 0 2.18 0 0.458 6.430 58.7 6.0622 3 222 18.7 394.12
## lstat  medv
## 1 4.98  24.0
## 2 9.14  21.6
## 3 4.03  34.7
## 4 2.94  33.4
## 5 5.33  36.2
## 6 5.21  28.7
```

For the sake of illustration, we will use the effect of variable *chas* as a random effect. The fixed effects we would like to consider are listed in the object *fxd* below.

```r
# make chas a factor to use it as a random effect
Boston$chas = as.factor(Boston$chas)
```

```r
rndm = c("(1 | chas)")
```

We would also like to keep the main effect of variable *rm* in our models, so we will include it in all models.

```r
include = c("rm")
```

Suppose that we are interested in the variable *crim*, per capita crime rate by town, and we would like to model the crime rate with the mixed model.
To find the most parsimonious model, we will start with the full model (1) and will use backward selection to drop the terms from the full model. We will use BIC as a discriminating criterion between different models. The input sinkit creates a file “results.txt” in R working directory and prints the output to this file that would, otherwise, be printed to the R console.

```r
# mixed model/backward/BIC
mod1 <- modelSelection(response = "crim", random = rndm, fixed = fxd, keep = include,
                        dat = Boston, direction = "backward", method = "BIC",
                        sinkit = "results.txt")
mod1
```

---

\[
\text{crim} \sim (1 \mid \text{chas}) + \text{tax} + \text{indus} + \text{nox} + \text{ptratio} + \text{black} + \text{medv} + I(\text{tax}^2) + I(\text{indus}^2) + I(\text{nox}^2) + I(\text{ptratio}^2) + I(\text{tax}^3) + I(\text{tax}^4) + \text{tax} : \text{indus} + \text{tax} : \text{nox} + \text{tax} : \text{rm} + \text{tax} : \text{ptratio} + \text{tax} : \text{black} + \text{tax} : \text{medv} + \text{indus} : \text{nox} + \text{indus} : \text{rm} + \text{indus} : \text{ptratio} + \text{rm} : \text{ptratio} + \text{tax} : \text{indus} : \text{nox} + \text{tax} : \text{rm} : \text{ptratio} + \text{rm} : \text{ptratio}
\]
The best model obtained through backward selection using BIC is given by

\[ \text{crim} \sim (1 \mid \text{chas}) + \text{tax} + \text{rm} + \text{medv} + I(\text{tax}^2) + I(\text{tax}^3) + \text{tax} : \text{medv} + I(\text{tax}^4) \]  

The BIC values for models (1) and (2) were 3446.42 and 3381.847, respectively. Model (2) was obtained in 11.69 seconds.

Here is a partial output that is stored in “results.txt” that informs the user about the selection process.

```
[1] "-----------------------------------------------------------"
[1] "-----------------------------------------------------------"
The model now has BIC value of 3446.422.
The model now has 26 terms, of which 25 are fixed effects and 1 are/is random effect(s).

There are 10 terms that are eligible to be dropped and these are:
```
[1] "tax:black"      "tax:medv"      "indus:rm"      "indus:ptratio"
[9] "I(ptratio^2)"  "I(tax^4)"
```

R is going to fit 10 model(s), each time removing one variable that is eligible to be dropped from the current model.

The 10 models were fitted in 2.03 seconds.

The top five alternative models with the lowest BIC are:

<table>
<thead>
<tr>
<th>Term Dropped</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>LRT</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>indus:rm</td>
<td>3326.123</td>
<td>3440.239</td>
<td>-1636.061</td>
<td>3272.123</td>
<td>0.8341859</td>
<td></td>
</tr>
<tr>
<td>tax:indus:nox</td>
<td>3326.147</td>
<td>3440.263</td>
<td>-1636.073</td>
<td>3272.147</td>
<td>0.7946782</td>
<td></td>
</tr>
<tr>
<td>tax:black</td>
<td>3326.212</td>
<td>3440.328</td>
<td>-1636.106</td>
<td>3272.212</td>
<td>0.7157352</td>
<td></td>
</tr>
<tr>
<td>I(ptratio^2)</td>
<td>3326.613</td>
<td>3440.729</td>
<td>-1636.306</td>
<td>3272.613</td>
<td>0.4651557</td>
<td></td>
</tr>
<tr>
<td>indus:ptratio</td>
<td>3327.127</td>
<td>3441.243</td>
<td>-1636.563</td>
<td>3273.127</td>
<td>0.3060152</td>
<td></td>
</tr>
</tbody>
</table>

The term that will be dropped is: indus:rm.

The updated model now has BIC value of 3440.239.

The updated model is now:

\[ \text{crim} \sim (1 \mid \text{chas}) + \text{tax} + \text{indus} + \text{nox} + \text{rm} + \text{ptratio} + \text{black} + \text{medv} + I(\text{tax}^2) + I(\text{indus}^2) + I(\text{nox}^2) + I(\text{ptratio}^2) + I(\text{tax}^3) + I(\text{tax}^4) + \text{tax:indus} + \text{tax:nox} + \text{tax:rm} + \text{tax:ptratio} + \text{tax:black} + \text{tax:medv} + \text{indus:nox} + \text{indus:ptratio} + \text{rm:ptratio} + \]
The model now has BIC value of 3440.239.

There are 9 terms that are eligible to be dropped and these are:

- tax:black
- tax:medv
- indus:ptratio
- tax:indus:nox
- tax:rm:ptratio
- I(indus^2)
- I(nox^2)
- I(ptratio^2)
- I(tax^4)

R is going to fit 9 model(s), each time removing one variable that is eligible to be dropped from the original model.

The 9 models were fitted in 0.54 secs.

The top five alternative models with the lowest BIC are:

<table>
<thead>
<tr>
<th>Term Dropped</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>LRT p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>tax:indus:nox</td>
<td>3324.191</td>
<td>3434.081</td>
<td>-1636.096</td>
<td>3272.191</td>
<td>0.7939316</td>
</tr>
<tr>
<td>tax:black</td>
<td>3324.248</td>
<td>3434.138</td>
<td>-1636.124</td>
<td>3272.248</td>
<td>0.7233357</td>
</tr>
<tr>
<td>I(ptratio^2)</td>
<td>3324.647</td>
<td>3434.537</td>
<td>-1636.323</td>
<td>3272.647</td>
<td>0.4691850</td>
</tr>
<tr>
<td>indus:ptratio</td>
<td>3325.178</td>
<td>3435.068</td>
<td>-1636.589</td>
<td>3273.178</td>
<td>0.3044141</td>
</tr>
<tr>
<td>I(indus^2)</td>
<td>3325.903</td>
<td>3435.792</td>
<td>-1636.951</td>
<td>3273.903</td>
<td>0.1822032</td>
</tr>
</tbody>
</table>

The term that will be dropped is: tax:indus:nox.

The updated model now has BIC value of 3434.081.

The updated model is now:

crim ~ (1 | chas) + tax + indus + nox + rm + ptratio + black +
    medv + I(tax^2) + I(indus^2) + I(nox^2) + I(ptratio^2) + 
    I(tax^3) + I(tax^4) + tax:indus + tax:nox + tax:rm + tax:ptratio + 
    tax:black + tax:medv + indus:nox + indus:ptratio + rm:ptratio + 
    tax:rm:ptratio

Example 2

This example uses the same data set, but it builds the final model using forward selection, where the models are compared with the likelihood ratio test with $\alpha = 0.05$. The procedure’s output is saved in results2.txt file and is not shown here.
Here is are the results of the forward selection procedure:

```
mod2
```

```
## $finalModel
## [1] "lmer(crim ~ (1 | chas) + rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4) + black)"
##
## $termsKept
## [1] "rm"
##
## $direction
## [1] "forward"
##
## $method
## [1] "LRT"
##
## $path
## LRT pValue
## c("tax:medv", "tax", "medv") 7.692058e-18
## c("I(tax^2)", "I(tax^3)", "I(tax^4)") 1.320418e-02
## black 4.425196e-02
##
## $alpha
## [1] 0.05
##
## $runningTime
## [1] "4.83 secs"
```

The output shows that, to the model containing \( \text{crim} \sim (1 \mid \text{chas}) \), in round one the algorithm added \( \text{tax:medv}, \text{tax}, \text{medv} \). Then, in round two, to the model \( \text{crim} \sim (1 \mid \text{chas}) + \text{tax:medv} + \text{tax} + \text{medv} \), the algorithm added \( I(\text{tax}^2), I(\text{tax}^3), I(\text{tax}^4) \). In round three, the algorithm added variable \( \text{black} \). In round four, the three way interaction \( \text{tax:rm:ptratio} \) could not be added since the LRT p-value \( \approx 0.91 > 0.05 \). The final model was then given by

\[
\text{crim} \sim (1 \mid \text{chas}) + \text{black} + \text{rm} + \text{tax} + \text{medv} + \text{tax:medv} + I(\text{tax}^2) + I(\text{tax}^3) + I(\text{tax}^4).
\]

**Example 3**

This example uses the same data set and it builds the final model using forward selection. Since there aren’t any random effects and we are potentially adding more than one term at a time (if an interaction is considered for inclusion, it is included with its main effects), the models are compared with an \( F \) test and \( \alpha = 0.025 \). The procedure’s output is not saved but is directly printed to the console.

```
# fixed model/forward/alpha = 0.025
mod3 <- modelSelection(response = "crim", fixed = fxd, 
                       direction = "forward", keep = include, dat = Boston, alpha = 0.025)
```

```
## [1] "-----------------------------------------------------------" 
## [1] "-----------------------------------------------------------"
## The model now has 2 terms, including the intercept, of which 1 are/is fixed effect(s).
##
## The groups of terms that are eligible to be added are:
```
## AddGroup
## 1 "tax:indus, tax, indus"
## 2 "tax:nox, tax, nox"
## 3 "tax:rm, tax"
## 4 "tax:ptratio, tax, ptratio"
## 5 "tax:black, tax, black"
## 6 "tax:medv, tax, medv"
## 7 "indus:nox, indus, nox"
## 8 "indus:rm, indus"
## 9 "indus:ptratio, indus, ptratio"
## 10 "rm:ptratio, ptratio"
## 12 "tax:rm:ptratio, tax:rm, tax:ptratio, rm:ptratio, tax, ptratio"
## 13 "tax, I(tax^2)"
## 14 "indus, I(indus^2)"
## 15 "nox, I(nox^2)"
## 16 "ptratio, I(ptratio^2)"
## 17 "tax, I(tax^2), I(tax^3)"
## 18 "tax, I(tax^2), I(tax^3), I(tax^4)"
## 19 "tax"
## 20 "indus"
## 21 "nox"
## 22 "ptratio"
## 23 "black"
## 24 "medv"
##
## The top five most significant group of terms are:
##
## | Terms Considered | F-test p-Value |
## |------------------|---------------|
## | tax:medv, tax, medv | 2.923040e-54 |
## | tax, I(tax^2), I(tax^3), I(tax^4) | 4.513501e-49 |
## | tax, I(tax^2) | 5.834653e-46 |
## | tax, I(tax^2), I(tax^3) | 3.925719e-45 |
## | tax:black, tax, black | 1.348082e-44 |
##
## The group of terms that will be added is: tax:medv + tax + medv.
##
## The updated model is now:
## crim ~ rm + tax:medv + tax + medv
## <environment: 0x7fd3681f67c0>
##
## The model now has 5 terms, including the intercept, of which 4 are/is fixed effect(s).
##
## The groups of terms that are eligible to be added are:
##
## AddGroup
## 1 "tax:indus, indus"
## 2 "tax:nox, nox"
## 3 "tax:rm"
## 4 "tax:ptratio, ptratio"
The top five most significant group of terms are:

1. \( I(tax^2), I(tax^3), I(tax^4) \)  
   \[ \text{Terms} \quad \text{Considered} \quad \text{F-test} \quad \text{p-Value} \]  
   \[ I(tax^2), I(tax^3), I(tax^4) \quad 3.184184e-08 \]

2. \( \text{tax:indus:nox, tax:indus, tax:nox, indus:nox, indus, nox} \)  
   \[ \text{Terms} \quad \text{Considered} \quad \text{F-test} \quad \text{p-Value} \]  
   \[ \text{tax:indus:nox, tax:indus, tax:nox, indus:nox, indus, nox} \quad 3.259510e-04 \]

3. \( I(tax^2) \)  
   \[ \text{Terms} \quad \text{Considered} \quad \text{F-test} \quad \text{p-Value} \]  
   \[ I(tax^2) \quad 9.077710e-03 \]

4. \( \text{indus, I(indus^2)} \)  
   \[ \text{Terms} \quad \text{Considered} \quad \text{F-test} \quad \text{p-Value} \]  
   \[ \text{indus, I(indus^2)} \quad 9.684604e-03 \]

5. \( \text{black} \)  
   \[ \text{Terms} \quad \text{Considered} \quad \text{F-test} \quad \text{p-Value} \]  
   \[ \text{black} \quad 1.380694e-02 \]

The group of terms that will be added is: \( I(tax^2) + I(tax^3) + I(tax^4) \).

The updated model is now:

\[ \text{crim} \sim \text{rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4)} \]

The model now has 8 terms, including the intercept, of which 7 are/is fixed effect(s).

The groups of terms that are eligible to be added are:

1. \( \text{tax:indus, indus} \)
2. \( \text{tax:nox, nox} \)
3. \( \text{tax:rm} \)
4. \( \text{tax:ptratio, ptratio} \)
5. \( \text{tax:black, black} \)
6. \( \text{indus:nox, indus, nox} \)
7. \( \text{indus:rm, indus} \)
8. \( \text{indus:ptratio, indus, ptratio} \)
9. \( \text{rm:ptratio, ptratio} \)
10. \( \text{tax:indus:nox, tax:indus, tax:nox, indus:nox, indus, nox} \)
11. \( \text{tax:rm:ptratio, tax:rm, tax:ptratio, rm:ptratio, ptratio} \)
12. \( \text{indus, I(indus^2)} \)
13. \( \text{nox, I(nox^2)} \)
## The top five most significant group of terms are:

<table>
<thead>
<tr>
<th>Terms Considered</th>
<th>F-test p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>black</td>
<td>0.04629174</td>
</tr>
<tr>
<td>tax:rm:ptratio, tax:rm, tax:ptratio, rm:ptratio, ptratio</td>
<td>0.08952398</td>
</tr>
<tr>
<td>tax:ptratio, ptratio</td>
<td>0.09601093</td>
</tr>
<tr>
<td>tax:black, black</td>
<td>0.13732558</td>
</tr>
<tr>
<td>tax:indus, indus</td>
<td>0.14690211</td>
</tr>
</tbody>
</table>

The group of terms that will be added is: black.

The updated model is now:
```
crim ~ rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4) +
black
```

Here are the results of this forward selection procedure:

```r
mod3
```

```r
$finalModel
[1] "lm(crim ~ rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4))"
```

```r
$termsKept
[1] "rm"
```

```r
$direction
[1] "forward"
```

```r
$path
TermsAdded       pValue
1         tax:medv, tax, medv 2.923040e-54
2       I(tax^2), I(tax^3), I(tax^4) 3.184184e-08
3         black 4.629174e-02
```

```r
$alpha
[1] 0.025
```

```r
$runningTime
[1] "0.47 secs"
```